

THIS WEEK

EDITORIALS

CANADA It is time to allow scientists to speak to the press **p.6**

WORLD VIEW The rights and wrongs of social neuroscience **p.7**



DISCOVERY Seven new species of limbless amphibians **p.8**

The darker side of stem cells

An investigation by Nature has found that patients in Texas are receiving unproven stem-cell treatments. The state and the company involved need to ensure that they follow FDA guidelines.

Stem cells offer the hope that one day they will be able to cure a huge range of disorders. But too many people are promising those cures to patients now, long before there is any evidence that they work. These claims are potentially misleading at best, and at worst could be downright harmful.

This week, *Nature* raises important questions about one company that works with adult stem cells: Celltex Therapeutics in Houston, Texas. *Nature's* investigation, reported on page 13, suggests that the company has supplied adult stem cells to Texas doctors who offer unproven treatments to patients, and that the company is involved in these treatments. One doctor claims that the treatments are part of a clinical study run by Celltex and that the company pays him US\$500 a time to inject the cells into patients, who are charged up to \$25,000 for a course. The US Food and Drug Administration (FDA) considers it to be a crime to inject unapproved adult stem cells into patients. David Eller, chief executive of Celltex, denies that the company is involved in treatment procedures, but would not comment on *Nature's* findings about how its cells are used or answer questions about them.

Celltex has the backing of state governor Rick Perry, who has tried adult-stem-cell treatments himself. And the company recently recruited Glenn McGee, editor-in-chief of the *American Journal of Bioethics*, to be its president of ethics and strategic initiatives. McGee, whose move from the academic world to industry sparked a controversy over a perceived conflict of interest (see *Nature* **482**, 449–450; 2012), promises that Celltex will set up and run clinical trials, and will do so according to strict ethical standards. But he too would not answer *Nature's* questions on whether the company knew about the unapproved clinical use of its stem cells or whether he considered such activity to be ethical and legal.

There is an ethical paradox here. How can Celltex propose clinical trials for stem-cell treatments while at the same time it is, according to a doctor involved, paying physicians to use those treatments — or supplying cells to doctors who would no doubt use them — in the clinic? Shouldn't clinical trials be done before a treatment is given to paying patients? Conversely, if a treatment is known to be safe and effective enough to be prescribed routinely, and for a sizeable fee, what is the point of doing a costly clinical trial?

The questions do not stop at Celltex. The governor-appointed Texas Medical Board is set to launch regulations in April to tighten controls on the use of 'investigational agents' — such as stem cells that do not have FDA approval. But the board is simultaneously offering an alternative route by proposing that those who want to use adult-stem-cell treatments need either FDA approval or simply the approval of a local institutional review board.

Texas officials should take the FDA's regulatory power over stem cells more seriously. In its discussions, the state's medical board revisited the well-trodden ground of whether adult-stem-cell procedures — in which cells are taken from a patient, processed and cultured, before being reintroduced — should be under FDA jurisdiction, or whether

they are akin to simple skin grafts from one part of the body to another, which do not require validation in an FDA-approved clinical trial.

The FDA can help to clarify these matters. A sensible first step in the new state regulations would be a requirement for any firm that plans to inject processed stem cells into patients to contact the FDA, which can

"Texas officials should take the FDA's regulatory power over stem cells more seriously."

advise on whether federal rules — the same federal rules that have already been used to arrest stem-cell practitioners and to stop a company pushing unapproved treatments elsewhere — apply to what they are doing. Once past that step, Texas could move on to develop its own safety regulations.

If the Texas Medical Board were to act according to its stated pledge to protect patients, then it would make clear the need for clinical validation of adult stem cells before use and would rescind the medical licences of any doctors in breach of rules on using unapproved treatments. If Celltex truly wants to help patients, then it should refuse to supply stem cells for medical procedures until those procedures are properly proven to be effective. And if the company is serious about demonstrating clinical effectiveness itself, then it should start by contacting the FDA about what needs to be done. ■

The great beyond

Progress on rare genetic diseases shows the medical value of outliers.

"Treasure your exceptions! When there are none, the work gets so dull that no one cares to carry it further. Keep them always uncovered and in sight. Exceptions are like the rough brickwork of a growing building which tells that there is more to come and shows where the next construction is to be." Geneticist William Bateson offered this advice in 1908, around the dawn of modern genetics following the rediscovery of Gregor Mendel's pea plant experiments, and it remains sound today.

Every empiricist must contend with exceptions to the rule, which can illuminate research in unpredictable ways. Bateson was urging vigilance in observing rare offspring of plants and animals, which may point to new phenomena that can inform us about the broader biological context. The same is true for rare diseases, which often have a genetic basis.

This being a leap year, the extra day, 29 February, was designated Rare Disease Day. It aims to give a voice to the families of millions of exceptional children who are born each year with a rare or undiagnosed disease. Liddle syndrome and Tangier disease, for example,

affect a tiny number of people, especially compared with high blood pressure, diabetes or heart disease. The attention these rare disorders receive from pharmaceutical programmes, the medical community and academic research has historically been correspondingly small, despite the significant disease burden they bring to the health-care systems of many countries.

Perhaps up to one-fifth of paediatric hospital admissions worldwide are the result of Mendelian disorders — uncommon diseases that follow inheritance patterns outlined by Mendel in the nineteenth century, usually caused by mutations in a single gene. If not deadly, these conditions often demand expensive life-long care.

Beyond the economic factors are the harrowing stories of parents trying to find treatment or even a diagnosis for a child with an exceptional problem. For rare diseases for which there are perhaps only one or two experts in the world, it typically takes a broad network of parents working together for the families to find help, or even information.

The rapid rise of genome sequencing and its increasing use in paediatric clinics offers some hope. Children with known diseases can be diagnosed faster than before, and those with previously unknown syndromes might be able to get a better understanding of the reason for their illness.

Rare disorders have conventionally been used to sharpen the tools of genetic medicine. Diseases that affect several families are prime targets for unpicking and understanding the effects of human gene mutations. With the development of better tools, researchers are now equipped to tackle even the rarest disorders. As the price of sequencing drops and analysis tools become more sophisticated,

fewer individuals are needed to pin down the common genetic cause.

As the News Feature on page 20 highlights, one small clinic serving Amish and Mennonite communities in the United States is at the forefront of these efforts. Through sequencing and other tools, the

“Genetic disease has a long history of teaching scientists about normal biological processes.”

paediatricians there estimate that they can discover 5–15 new disorders per year. They think that about half should be treatable if the underlying cause is identified early.

The clinic is in an exceptional position, however. The communities it serves come from small founder populations, have a good knowledge of family history and a high rate of intermarriage. All of this makes unravelling the genetics of disease easier than

in a more diverse population. But there are two reasons to treasure these exceptions. First, genetic disease has a long history of teaching scientists about normal biological processes, and can sometimes give insight into the processes that go awry in other, more common diseases. Liddle syndrome, for example, is one of several genetic disorders in which high blood pressure is a major defining factor. Tangier disease causes low levels of ‘good’ cholesterol in the blood, and many of those affected have premature atherosclerosis.

The second reason relates to a growing appreciation of the heterogeneity of ‘common’ diseases. Projects that sequence tumour genomes from dozens of patients — who by standard diagnostic measures have the same cancer, and thus the same disease — have revealed that each tumour has unique and divergent genetic properties. Indeed, the real work may begin once we realize that every case is an exception. ■

Frozen out

Canada’s government should free its scientists to speak to the press, as its US counterpart has.

Media interactions with government scientists have undergone a reversal across North America during the past six years. In the United States, President Barack Obama’s administration has directed federal science agencies to develop integrity policies with clear guidelines for scientists who are approached by journalists.

In December, agencies including the US National Science Foundation (NSF) and the US National Oceanic and Atmospheric Administration (NOAA) issued guidelines that promote openness with the press. For instance, NOAA and NSF-funded scientists and staff are free to speak to journalists without first seeking the approval of a public-affairs officer. The NSF’s policy states that researchers are free to express their personal views as long as they make clear that they are not speaking on behalf of the agency. And scientists also have right of review over agency publications and press releases that claim to represent their expert opinions. Such policies may not be implemented successfully in all cases, but they show that attitudes have evolved encouragingly since 2006, when charges that then-president George W. Bush’s administration had silenced US government researchers made front-page news.

Over the same period, Canada has moved in the opposite direction. Since Prime Minister Stephen Harper’s Conservative Party won power in 2006, there has been a gradual tightening of media protocols for federal scientists and other government workers. Researchers who once would have felt comfortable responding freely and promptly to journalists are now required to direct inquiries to a media-relations office, which demands written questions in advance, and might not permit scientists to speak. Canadian journalists have documented several instances in which prominent researchers have been prevented

from discussing published, peer-reviewed literature. Policy directives and e-mails obtained from the government through freedom of information reveal a confused and Byzantine approach to the press, prioritizing message control and showing little understanding of the importance of the free flow of scientific knowledge.

The Harper government’s poor record on openness has been raised by this publication before (see K. O’Hara *Nature* **467**, 501; 2010), and *Nature*’s news reporters, who have an obvious interest in access to scientific information and expert opinion, have experienced directly the cumbersome approval process that stalls or prevents meaningful contact with Canada’s publicly funded scientists. Little has changed in the past two years: rather than address the matter, the Canadian government seems inclined to stick with its restrictive course and ride out all objections.

That position is coming under increasing pressure as a result of the scientific-integrity policies taking shape across the border. The clarity of the US guidelines undercuts the Canadian government’s assertion that its own media policies are adequate and have simply been misunderstood. If the Harper government truly embraces public access to publicly funded scientific expertise, then it should do what the Canadian Science Writers’ Association and several other organizations have called for in a letter sent to the prime minister on 16 February: “implement a policy of timely and transparent communication” like those used by NOAA and the NSF.

The letter coincided with a symposium, ‘Unmuzzling Government Scientists: How to Re-open the Debate’, which was held last week at the meeting of the American Association for the Advancement of Science in Vancouver, Canada. With the country taking centre stage as the meeting’s host, the Harper government found its media policies in the international spotlight. Scientists and other visitors from around the globe discovered, to their surprise, that

Canada’s generally positive foreign reputation as a progressive, scientific nation masks some startlingly poor behaviour. The way forward is clear: it is time for the Canadian government to set its scientists free. ■

➔ **NATURE.COM**

To comment online,
click on Editorials at:
go.nature.com/xhnuqv

F. BONNEFOI



Clear up this fuzzy thinking on brain scans

France has banned commercial applications of brain imaging. So why approve its use in court, asks Olivier Oullier.

Neuroscience could not save Edwin Hart Turner. After years on death row, the double murderer was executed by lethal injection at the Mississippi State Penitentiary in Parchman on 8 February, despite last-minute legal appeals including brain scans that, his defence claimed, showed that he was mentally ill and therefore unfit for the death penalty. Whether or not we neuroscientists like it, our field has become the new genetics, the latest scientific field to be used and sometimes hijacked to explain human behaviour.

Hardly a week goes by without the media showcasing beautiful three-dimensional images of the brain in action, which supposedly explain how and why humans do the things we do. Most recently, people have pretended that they can use neuroimaging alone to identify paedophiles or prove that we fall in love with mobile phones.

The limits of the technology, together with our incomplete picture of how the brain functions, make it hard to take these claims seriously. But, regardless of the doubts expressed by many neuroscientists, the appeal of brain images, and the illusion they offer of understanding and explaining our daily behaviour, continues to grow.

The spreading use of neuroscience outside research and medical labs raises ethical and practical questions. Commercial services, such as marketing and lie detection, are already starting to make use of brain-imaging techniques. These services are over-interpreting the science. Although neuroimaging could help us to understand how people make decisions, it should not be sold as something that can predict or judge human behaviour. A brain on its own tells us nothing.

That a technique is still in its infancy doesn't stop people from trying to sell it or from buying it. And damage can be done even if the victims of neuromarketing hype are not the general public but the gullible heads of companies who are being overcharged.

If this business expands, it will become the most visible face of neuroimaging. We cannot afford to have public opinion turned against the development of neuroimaging because of overstated claims by commercial opportunists.

France has tried to crack down on such rogue uses of neuroscience. With the help of myself and other neuroscientists, the French parliament has revised its 2004 rules on bioethics. The result, passed last year, is a section of the law that simply states: "Brain-imaging methods can be used only for medical or scientific research purposes or in the context of court expertise." The revised law effectively bans the commercial use of neuroimaging in France, although neuromarketing companies have only to cross the border to continue their business.

The ban was controversial from the start, with many experts arguing that neuroimaging should not be singled out. Why draw attention to bogus commercial uses of the technology, they said. This was a somewhat hypocritical stance, given that these scientists consider such applications to be unscientific by default, yet they argued that neuroimaging can improve understanding of brain disorders and of how children learn to read, for example. That attitude would leave us in the undesirable situation in which the reliability of an innovative methodology is judged on its perceived value and utility for society, not on the science itself.

In view of the uncertainties in the technique, the revised law's invitation to use neuroimaging in courts is also problematic — certainly, none of the neuroscientists consulted during the drafting process said that it should be encouraged.

So far, neuroimaging has been confined mostly to a supporting role in court: in sentence mitigation, for instance. But there have been misguided and dangerous attempts in India, Italy and the United States to use brain scans as key and decisive evidence of guilt or innocence. Such efforts have not been successful, but we get closer every day to the situation that a court somewhere in the world will accept neuroimaging data as primary evidence. France should not be encouraging its use at this stage, at least not in such nonspecific terms.

French politicians' call for neuroimaging to be used in courts even though no expert advocates the move speaks volumes about the excessive trust they have in this emerging technology. Perhaps we are seeing the consequences of the hype that surrounded the early studies and the fantasies promoted by companies who profit from the technology. Maybe this excitement, along with attempts by academic neuroscientists to interest policy-makers in the field, helped to convince the politicians that, although the brain sciences alone will not provide definitive answers to societal issues, it would be a mistake to ignore their insight and potential. We should support efforts in that direction, but is there yet enough evidence to give the green light to neuroimaging in the courts? Certainly not.

Brain scientists may not be oracles, but our research, responsibly interpreted, can help policy-makers to make informed decisions. As such, it should be given the opportunity to progress. Law and science have something in common — both can be misinterpreted. ■

THAT A TECHNIQUE
IS STILL IN ITS
INFANCY
DOESN'T STOP
PEOPLE FROM
TRYING TO
SELL IT
OR FROM BUYING IT.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/me4gmm

Olivier Oullier is at Aix-Marseille University, a Young Global Leader and member of the Global Agenda Council on Brain Cognitive Sciences for the World Economic Forum and adviser to the Center for Strategic Analysis of the French Prime Minister.
e-mail: olivier@oullier.fr

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTROPHYSICS

Rare element from space

The element tellurium probably originated in exploding stars, or supernovae, through a process by which nuclei rapidly capture neutrons and become heavier.

Ian Roederer of the Carnegie Observatories in Pasadena, California, and his colleagues used the Hubble Space Telescope to detect signs of tellurium in three ancient stars in the halo of the Milky Way. Tellurium is the heaviest element for which production by rapid neutron capture, called the *r*-process, can be predicted using laboratory data.

The authors report that the observed levels of tellurium match earlier predictions made for the *r*-process in the Solar System. They conclude that tellurium and other heavy elements are produced predominantly by this process.

Astrophys. J. 747, L8 (2012)

BIODIVERSITY

New amphibians with no limbs

A family of limbless amphibians has been discovered in the soils of northeastern India.

Sathyabhama Das Biju at the University of Delhi and his colleagues unearthed more than 500 examples

encompassing seven new species of the caecilian order of amphibians. An analysis of the animals' genomes and cranial anatomy showed that these Chikilidae (pictured with a clutch of eggs) — which measure up to 25 centimetres in length — probably first appeared during the early Cretaceous period, about 140 million years ago.

The animals' closest known relatives are in Africa. The authors suggest that, historically, caecilians are likely to have been geographically restricted, with low rates of speciation.



CLIMATE SCIENCE

The extended reach of Australian drought

The Big Dry, a prolonged drought that affected southeast Australia from 1997 to 2011, was more extensive than previously thought.

Gavan McGrath at the University of Western Australia in Crawley and his colleagues analysed satellite data from across the continent and found evidence of decreased water storage, rainfall and plant growth throughout the country between 2002 and 2010. In the southeast, the drought correlated with an irregular Indian Ocean circulation,

whereas in the northwest it was associated with a decreased frequency of tropical cyclones. The authors say that the northwest drought coincided with and probably exacerbated the one in the southeast.

The findings suggest that distinct climatic factors such as decadal cyclone trends and changes in ocean circulation can combine to create a continental-scale drought.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2011GL050263> (2012)

EVOLUTION

Invasive mosquito adapts fast

Since arriving in the United States from Japan in 1985, the invasive Asian tiger mosquito has spread across the country's eastern reaches, providing a natural demonstration of

adaptive evolution. Peter Armbruster at Georgetown University in Washington DC and his colleagues report that over roughly 20 years, the insect (*Aedes albopictus*) has adapted to differences in the timing of winter onset across roughly 15° of latitude.

The authors measured the day length required to induce winter-time dormancy in mosquitoes across various latitudes in the US and Japan between 2005 and 2008, and compared the measurements with those taken about 20 years earlier. They found a response to seasonal change that is among the fastest documented

A. TOENSING/GETTY

S. D. BIJU

adaptation in nature.

Other traits, such as body size, did not show a pattern of adaptation in either country. The findings support the idea that, in a changing climate, animals evolve mainly by adapting to changes in seasonal shifts.

Am. Nat. <http://dx.doi.org/10.1086/664709> (2012)

MICROBIOLOGY

New species spring forth

Microorganisms in a Siberian hot spring have been caught in the act of diverging to form two species.

Rachel Whitaker at the University of Illinois in Urbana-Champaign and her team sequenced the genomes of 12 strains of *Sulfolobus islandicus* — an archaeon living in hot springs around a volcano on Russia's Kamchatka Peninsula. The sequences show that the microbes belong to two groups — dubbed red and blue — and that genes are exchanged more readily within than between the groups. This means that each group of microbes meets the classic definition of a biological species.

The authors think that physiological differences between the two populations could be preventing gene flow. For instance, strains of the red group replicate faster and to higher densities than do those of the blue group.

PLoS Biol. 10, e1001265 (2012)

GEOLOGY

Stretch marks on the Moon

The Moon has been stretched within the past 50 million years — a surprising and relatively recent sign of extensional tectonics for a body that has been around for 4.5 billion years.

Using a camera on NASA's

Lunar Reconnaissance Orbiter, Thomas Watters of the Smithsonian Institution in Washington DC and his colleagues spotted graben — long, narrow blocks of rock that drop down and form trenches as the Moon's crust is stretched. Some of the graben are as shallow as one metre, suggesting that, in geological terms, they are fresh.

The authors suggest that their findings are inconsistent with models that predict that the primordial Moon

was completely molten and would have contracted so much that local extensional pressures would have been quashed. Instead, the graben may reflect an environment of

weaker contractions that resulted from the early Moon having only a molten exterior. **Nature Geosci.** <http://dx.doi.org/10.1038/ngeo1387> (2012)

NEUROSCIENCE

Cognitive boost to brain connections

By sensitizing the process by which neurons normally alter the strength of their connections, a molecule derived from a neuronal protein improves learning and memory in rats.

José Esteban at the Autonomous University of Madrid and his colleagues studied the effects of the molecule, FGL, on rats and on slices of the rat hippocampus, a brain region involved in learning and memory. They showed that FGL caused persistent activation of signalling molecules in hippocampal neurons. This set in motion a chain of molecular events that increased the efficiency with which a type of receptor called AMPA was inserted into neuronal connections, or synapses. The incorporation of additional AMPA receptors into synapses

COMMUNITY CHOICE

The most viewed papers in science

MOLECULAR BIOLOGY

Noncoding RNAs decapped

HIGHLY READ
on www.cell.com
22 Jan–21 Feb

Long RNA molecules have recently been discovered that seem to regulate genes rather than code for proteins, but little is known about how they do this. Loss of a protective cap from these long noncoding RNAs (lncRNAs) seems to be a crucial step in regulating a specific type of gene in yeast.

Like protein-encoding messenger RNAs, lncRNAs have a cap that protects the molecule from decomposing. By preventing loss of the cap in budding yeast, Jeff Collier and his colleagues at Case Western Reserve University in Cleveland, Ohio, found that many lncRNAs function at genomic regions near highly regulated genes that respond to specific environmental cues such as sugars and iron. One family of these 'inducible' genes — the GAL system, which enables cells to metabolize a sugar called galactose — is regulated by lncRNAs that must lose their caps for the cell to activate the sugar-processing genes.

Mol. Cell 45, 279–291 (2012)

is known to improve synaptic plasticity.

The authors hope that FGL could provide a starting point for the development of drugs that boost cognition.

PLoS Biol. 10, e1001262 (2012)

PALAEOECOLOGY

Ancient forest preserved in ash

Volcanic ash buried a swampy forest roughly 298 million years ago, preserving a wealth of detail about the region's flora. Scientists have uncovered the fossilized plants in what is now northern China.

Jun Wang at the Chinese Academy of Sciences in Nanjing, Hermann Pfefferkorn at the University of Pennsylvania in Philadelphia and their team reconstructed the ancient ecosystem by analysing the positions of individual plants across three sites (an artist's impression pictured) that cover a combined area of more than 1,000 square metres. Besides sporting a broad, low canopy of tree ferns, the peat forest contained trees that



looked like feather dusters and reached heights of 25 metres or more. The team also found fossils of vines and three species of Noeggerathiales — small spore-bearing trees thought to be close relatives of the earliest ferns.

Proc. Natl Acad. Sci USA <http://dx.doi.org/10.1073/pnas.1115076109> (2012)

For a longer story on this research, see go.nature.com/eklwre

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Oceans partnership

A coalition headed by the World Bank hopes to raise US\$1.5 billion to help to improve the world's oceans. Unveiling the Global Partnership for Oceans in Singapore on 24 February, the bank's president, Robert Zoellick, said that governments, non-governmental organizations, private companies and United Nations bodies had joined to coordinate their approaches to ocean conservation. Many of the partnership's targets — such as rebuilding depleted fish stocks and increasing the number of marine protected areas — mirror existing UN goals. See go.nature.com/he6ldt for more.

Alzheimer's plan

The US Department of Health and Human Services has launched a draft action plan to find ways to treat and prevent Alzheimer's disease by 2025.

The plan is part of a concerted effort focusing on Alzheimer's research and care; last month, President Barack Obama's administration told the US National Institutes of Health to reallocate US\$50 million for the disease from its 2012 budget (boosting its Alzheimer's research funding to \$498 million), and requested an extra \$80 million for Alzheimer's research in 2013.

Greek protests

Demonstrations continued last week against reforms that aim to introduce more meritocracy and accountability into Greece's higher-education system. On 22 February, protesters disrupted elections for governing boards at the University of Crete in Heraklion and at the Athens University of Economics and Business. Elections have



ARMADA DE CHILE

Fire destroys Brazil's Antarctic outpost

Brazil's scientific community is reeling from a fire that blazed through the country's Comandante Ferraz research station (pictured) in the South Shetland Islands off the Antarctic Peninsula, killing two navy personnel and injuring another. The cause was unknown as

Nature went to press, but the Brazilian navy said that the fire began on 25 February and destroyed about 70% of the facility. Brazil's president, Dilma Rousseff, has vowed to rebuild the base and continue its environmental and biological research. See go.nature.com/iemahf for more.

been stopped at four other universities in past weeks, and more are scheduled for March. See page 15 for more.

Publish or perish

Cost-saving cuts at the University of Sydney, Australia, have triggered a furious backlash, after the university sent letters to around 100 academics last week, sacking them for not publishing frequently enough. Researchers were told that they had been fired because they had not published at least four "research outputs" over the past three years, says Michael Thomson, president of the Sydney branch of the Australian National Tertiary Education Union. The union was disputing the sackings as *Nature* went to press. See go.nature.com/rohppo for more.

Tar-sands deadlock

European diplomats have failed to agree on a measure that would effectively block imports of oil from tar sands in Canada. A vote on 23 February over whether to classify tar-sands oil as highly polluting in terms of greenhouse-gas emissions ended in stalemate; environment ministers are now expected to make a decision in June. Europe's fuel directive requires fuel suppliers to cut emissions, so a high emissions rating would in practice shut down the European market for tar-sands oil. See go.nature.com/sejma4 for more.

Science and Technology near Vienna, has secured its long-term future with a funding deal worth up to €990 million (US\$1.3 billion). On 22 February, the institute announced that its federal and provincial government co-funders would grant €65.8 million annually between 2017 and 2026, with another €33.2 million available for reaching agreed research-quality criteria, and for matching external grants won by its scientists. The Austrian Academy of Sciences called the funding "an unjustified privilege". See go.nature.com/qiciw5 for more.

FUNDING

Austrian largesse

Austria's most ambitious research institute yet, the three-year-old Institute of

Africa eco-network

In an effort to improve agricultural systems in Africa, researchers in the United States and South Africa have launched an initiative

to gather and integrate data about ecosystems, farming practices and human well-being across the continent. Backed by a US\$10-million grant from the Bill & Melinda Gates Foundation, the three-year project is expected to set the stage for a larger global monitoring system, organizers announced on 23 February at a meeting of the International Fund for Agricultural Development in Rome. See go.nature.com/bbhbm3 for more.

RESEARCH

Too quick, too soon

Two possible sources of error have been identified in the OPERA experiment, which last September announced that it had timed neutrinos travelling faster than the speed of light. On 22 February, the multinational collaboration announced that the errors related to the Global Positioning System signal used to synchronize the atomic clocks used in the experiment. See page 17 for more.

PEOPLE

Heartland fallout

The Pacific Institute in Oakland, California, is investigating its president, Peter Gleick (**pictured**), after he admitted that he had dishonestly acquired,



and leaked to the media, internal budget documents from the Heartland Institute, a libertarian think tank in Chicago, Illinois, that aims to combat climate science. On 24 February, Gleick asked for a temporary leave of absence from the Pacific Institute, which he co-founded and has led for 25 years. On the same day, US Congressman Edward Markey (Democrat, Massachusetts) asked the Heartland Institute to confirm which of the documents were authentic.

Plagiarism politics

C. N. R. Rao, the science adviser to India's prime minister, has found himself embroiled in a plagiarism row, after he failed to notice some copied text in a paper that he had co-authored (B. Chitara *et al. Adv. Mater.* **23**, 5419–5424; 2011). Three sentences in the introduction and a

description of an equation had been copied verbatim from a referenced source. A PhD student, Basant Chitara, says he was responsible for the error. Rao says that the Indian media took a subsequent apology from all the authors to mean that the entire paper had been plagiarized. See go.nature.com/wbckgc for more.

BUSINESS

Obesity pill

An advisory panel to the US Food and Drug Administration (FDA) has recommended approving a diet pill that was rejected two years ago because of concerns over its side effects. Qnexa (phentermine plus topiramate) is made by Vivus of Mountain View, California, which saw its share price soar 78% when the drug received the thumbs-up on 23 February. The FDA is expected to decide on Qnexa by 17 April. If approved, it would be the first new US obesity drug in more than a decade.

Cancer lawsuits

Cancer geneticist Craig Thompson, head of the Memorial Sloan-Kettering Cancer Center in New York, is being sued by his former employers. The University of Pennsylvania in Philadelphia says that

COMING UP

5 MARCH

In New Orleans, Louisiana, a trial begins over who should pay for damages caused by the 2010 oil spill in the Gulf of Mexico. BP may settle some of its costs in advance.

7–8 MARCH

The European Space Agency hosts a symposium in Frascati, Italy, on international security — including threats from computer viruses and cyberwarfare.

go.nature.com/raj6a5

Thompson did not disclose research to the university that he later used to form a commercial biotechnology firm, Agios Pharmaceuticals in Cambridge, Massachusetts. The lawsuit, announced on 22 February, is the second against Thompson and Agios: last December, the Abramson Family Cancer Research Institute at the University of Pennsylvania — which Thompson previously headed — also sued him for not notifying university officials of his work. Both suits seek damages for the loss of intellectual property.

GM energy crops

Ceres, a US biotechnology firm that selectively breeds and genetically modifies seeds to grow high-yielding biofuel crops, raised US\$65 million in its initial public offering on 22 February. The firm, based in Thousand Oaks, California, sells sorghum seeds — used to make ethanol — in Brazil and sorghum and switchgrass seeds in the United States. It is 5% owned by the transgenic-seed giant Monsanto, of St Louis, Missouri.

► **NATURE.COM**

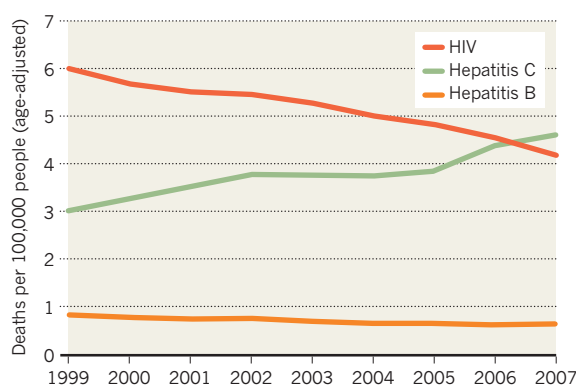
For daily news updates see:
www.nature.com/news

TREND WATCH

Two-thirds of the 3.2 million people in the United States living with the hepatitis C virus (HCV) were born between 1945 and 1964. These middle-aged 'baby-boomers' are now reaching the age when HCV-related diseases and deaths start to show up. A study documenting how HCV has overtaken HIV as a cause of death (see chart; from K. N. Ly *et al. Ann. Intern. Med.* **156**, 271–278; 2012) says that better awareness and treatments are needed.

RIISING DEATHS FROM HEPATITIS C

By 2007, hepatitis C infection had surpassed HIV as a cause of death in the United States, a new analysis reveals.



NEWS IN FOCUS

GREECE Campus protesters rally to preserve student power **p.15**

REPRODUCTION Egg-making stem cells might buy time for women **p.16**

FUNDING Budget threatens health of US disease agency **p.19**



PROFILE Michael Merzenich, brain-game guru, plays a new hand **p.24**

TYLER RUDICK



Celltex hosts the largest stem-cell bank in the United States.

REGENERATIVE MEDICINE

Stem-cell therapy takes off in Texas

A boom in unproven procedures is worrying scientists.

BY DAVID CYRANOSKI

With Texas pouring millions of dollars into developing adult stem-cell treatments, doctors there are already injecting paying customers with unproven preparations, supplied by an ambitious new company.

The US Food and Drug Administration (FDA) has not approved any such stem-cell treatment for routine clinical use, although

it does sanction them for patients enrolled in registered clinical trials. Some advocates of the treatments argue, however, that preparations based on a patient's own cells should not be classed as drugs, and should not therefore fall under the FDA's jurisdiction.

There are certainly plenty of people eager to have the treatments. Texas governor Rick Perry, for instance, has had stem-cell injections to treat a back complaint¹, and has supported legislation to help create banks to

store patients' harvested stem cells.

One company that has benefited from this buoyant climate is Celltex Therapeutics, which "multiplies and banks" stem cells derived from people's abdominal fat, according to chairman and chief executive David Eller. Its facility in Sugar Land, just outside Houston, opened in December 2011 and houses the largest stem-cell bank in the United States.

Celltex was founded by Eller and Stanley Jones, the orthopaedic surgeon who performed Perry's procedure, and it uses technology licensed from RNL Bio in Seoul. Because clinical use of adult-stem-cell treatments are illegal in South Korea, RNL has since 2006 sent more than 10,000 patients to clinics in Japan and China to receive injections.

Celltex says that although it processes and banks cells, it does not carry out stem-cell injections. It declined to answer *Nature's* questions about whether its cells have been used in patients. But there is evidence that the company is involved in the clinical use of the cells on US soil, which the FDA has viewed as illegal in other cases.

PUBLIC HYPE

In addition to the publicity surrounding Perry's treatment, a woman named Debbie Bertrand has been blogging about her experiences during a five-injection treatment with cells prepared at Celltex. Her blog (<http://debbiebertrand.blogspot.com>) hosts photographs of herself alongside Jones; Jennifer Novak, a Celltex nurse; Jeong Chan Ra, chief executive of RNL Bio; and her doctor, Jamshid Lotfi, a neurologist who works for the United Neurology clinic in Houston. Another photo is captioned: "My cells are being processed in here for my next infusion!!!" A third shows Bertrand, Lotfi and a physician called Matthew Daneshmand, who is, according to the caption, injecting Bertrand's stem cells into an intravenous drip, ready for the infusion. *Nature* has been unable to contact Bertrand.

Lotfi says that he has administered cells processed by Celltex to more than 20 people. "Five or six" — including Bertrand — have multiple sclerosis and "four or five" have Parkinson's disease, he says. Lotfi explains that patients sign up for treatment by contacting Novak, and that cells are prepared by removing about five grams of fat — containing roughly 100,000 mesenchymal stem

► cells — from the patient's abdomen. Over a three-week period, the cells are cultured until they reach about 800 million cells. Lotfi says that patients get at least three injections of 200 million cells each, and that the cells do not take effect for a few months. According to Lotfi, Celltex charges US\$7,000 per 200 million cells, and pays Lotfi \$500 per injection.

Lotfi says that most of his patients claim to get better after the treatment, but he admits that there is no scientific evidence that the cells are effective. "The scientific mind is not convinced by anecdotal evidence," he acknowledges. "You need a controlled, double-blind study. But for many treatments, that's not possible. It would take years, and some patients don't have years."

"The worst-case scenario is that it won't work," he adds. "But it could be a panacea, from cosmetics to cancer." He says that Celltex is conducting a trial in which patients "will be their own control". "If you can compare before and after and show improvement, there's no need for a placebo," he explains. "How can you charge people, and then give them a placebo?"

STANDARD PRACTICE

The answer, according to most biomedical researchers, is to follow the normal strategy for an unproven medicine: conduct a placebo-controlled clinical trial in which patients are not charged. "This is crucial for clinical trials to yield meaningful data and to avoid the placebo effect," says Joshua Hare, director of the Interdisciplinary Stem Cell Institute at the University of Miami in Florida, who worked with the biotechnology company Osiris Therapeutics in Columbia, Maryland, to conduct a trial of adult mesenchymal stem cells in patients who have had a heart attack².

The FDA does exempt some "compassionate use" procedures from these requirements, but only for very small numbers of patients with life-threatening disorders that demand immediate intervention.

Eller says that Celltex has "initiated a programme of clinical research" that has been approved by an independent review board in Texas and is designed to meet the standards set by the International Society for Stem Cell Research (ISSCR) in Deerfield, Illinois. "Our clinical trials are led by physician investigators who have completed training in both the

use of autologous mesenchymal stem cells and in research ethics," he explains. But he would not say whether the trials had started or been approved by the FDA, nor who would run them, what disorders would be targeted or whether patients would be charged. Meanwhile, RNL has published a phase I trial showing that its treatment for spinal-cord injury is safe³, and plans a phase II trial in South Korea to establish efficacy. It has also applied to the FDA to conduct a clinical trial in the United States.

Stem-cell scientists contacted by *Nature* are concerned by Celltex's activity. "Because we know so little about mesenchymal stem cells and whether they are indeed effective for

"The scientific mind is not convinced by anecdotal evidence."

treating any condition, I'd be very wary of how they are being infused into patients, and certainly concerned if practitioners are charging patients for medical procedures that haven't been proven to work and could in fact be harmful," says George Daley, director of the Stem Cell Transplantation Program at Harvard Medical School in Boston, Massachusetts, who helped the ISSCR to draft its guidelines. In the opinion of Arnold Kriegstein, director of the Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research at the University of California, San Francisco, "the very nature of Celltex's business plan, which involves charging patients considerable fees for so-called treatments for diseases and disorders for which there is no good clinical evidence of efficacy, crosses an ethical line".

Celltex recently brought in Glenn McGee, the editor-in-chief of the *American Journal of Bioethics*, to be its first president for ethics and strategic initiatives⁴. "I am working to build a programme of Celltex-run clinical trials that exceed the ethical standards promulgated by the best document to date on the matter, that from the ISSCR," says McGee. However, like Eller, McGee declined to answer *Nature's* questions about the company's knowledge of or involvement in the clinical use of its stem cells, nor would he talk about the legality and ethical nature of such practices.

The legal standing of stem-cell treatments is currently being debated in a court

case brought by Regenerative Sciences of Broomfield, Colorado, which was ordered by the FDA in 2010 to stop administering mesenchymal stem cells to patients⁵. One of the key issues being debated is whether the cells are "minimally manipulated" before being reinjected into the patient. Treatment with the patient's own, unprocessed tissue does not always require FDA approval.

DEGREE OF MANIPULATION

In a previous version of his LinkedIn profile, McGee asserted that Celltex "minimally modified" its cells. But Rita Chappelle, a spokeswoman for the FDA's Center for Biologics Evaluation and Research in Silver Spring, Maryland, contends that any "expanded" cells cannot be considered minimally manipulated. Another source, who used to work for the FDA's cellular therapy division and requested anonymity because they now work in the private sector, says that any cells kept outside the body overnight before injection are considered more than minimally manipulated.

The Texas Medical Board, which regulates physicians in the state, has now stepped into the fray. On 10 February, it approved draft rules that will require physicians to gain approval from an independent review committee before treating patients with adult stem cells. This stipulation, expected to become law in April, is meant to ensure patient safety and give physicians "a reasonable and responsible degree of latitude in the kinds of therapies they offer their patients", the draft rule says.

But Douglas Sipp, an expert in stem-cell regulations and ethics at the RIKEN Center for Developmental Biology in Kobe, Japan, believes that the Texas Medical Board is merely delegating the regulation of adult-stem-cell treatments to local review boards, which have little accountability, when the responsibility should lie with the FDA.

It is also unclear how Celltex will be affected by the rules. Lotfi, for one, does not think that they will hamper his treatments greatly. "It will be just a little more red tape," he says. ■ **SEE EDITORIAL P.5**

1. Cyranoski, D. *Nature* **477**, 377–378 (2011).
2. Hare, J. H. et al. *J. Am. Coll. Cardiol.* **54**, 2277–2286 (2009).
3. Ra, J. C. et al. *Stem Cells Dev.* **20**, 1297–1308 (2011).
4. Cyranoski, D. *Nature* **482**, 449–450 (2012).
5. Cyranoski, D. *Nature* **466**, 909 (2010).

 **MORE ONLINE**

TOP STORY



Scientists report back from Fukushima exclusion zone go.nature.com/w3gyvt

MORE NEWS

- Optical memory could ease Internet bottlenecks go.nature.com/l1buga
- Researchers chart intellectual trends in arXiv go.nature.com/uq11tl
- Budget cuts force lay-offs at UN biotechnology centre go.nature.com/h7wrnp

PODCAST



Patterns of human movement, exploring a fossil forest, and who really deciphered hieroglyphs? go.nature.com/w5vqad



Student protesters in Thessaloniki object to changes in university governance.

HIGHER EDUCATION

Protests delay Greek university reform

Small groups of agitators stop elections for governing boards.

BY ALISON ABBOTT

As European leaders put the finishing touches to a €130-billion (US\$170-billion) rescue package for Greece's enfeebled economy, angry demonstrators are blocking university reforms that many researchers believe are crucial to the country's recovery. The protesters, mostly students but also some academics, are targeting a law passed by the Greek parliament last August that seeks to introduce more meritocracy, dynamism and accountability to Greece's rigid

higher-education system. Opponents dislike the law mainly because it takes away students' rights to vote on many faculty decisions, and mandates 15-person governing boards that will have to include six members external to the university (see *Nature* **481**, 123–124; 2012).

On 22 February, demonstrators at the University of Crete in Heraklion and at the Athens University of Economics and Business disrupted elections for the governing boards by blocking the entrances to the buildings in which voting was to take place. Protesters have stopped elections at four other universities in past weeks.

The actions have caused widespread consternation. "It's frustrating," says pharmacologist Achilles Gravanis of the University of Crete, a member of the government's scientific advisory committee. "If we can't implement a law that had been approved by an overwhelming parliamentary majority, how are we going to convince those in Europe who are bailing us out that we are capable of reform?"

Greek higher-education minister Anna Diamantopoulou says that the protesters are "holding hostage the vast majority of faculty and students". She plans to ask parliament for rapid approval of an electronic voting system to bypass the confrontations before elections at other universities — eight of which are planned for March. Nine universities have not started planning elections because their rectors disapprove of the law.

Like other opponents, Yannis Krestenitis, an oceanographer at the Aristotle University of Thessaloniki, hopes that the law will be declared unconstitutional by the Supreme Court, which is deliberating charges that it violates academic independence. "It is clearly undemocratic," he says. Krestenitis helped to lead demonstrations that blocked two attempts to hold elections at the university, on 15 and 17 February.

Many others, however, regard the protests themselves as undemocratic. "What we have been seeing in Thessaloniki runs counter to any rules of a democratic state or decently functioning university," says physicist Orestis Kalogirou, a candidate for the Aristotle University board. According to Georgios Theodoridis, an analytical chemist at the institution, students broke into the polling station and were later joined by more than 150 protesters — including 20 or so faculty members — who formed a human chain to block the main university entrance while chanting slogans. "It feels a bit like George Orwell's *Animal Farm*, where the dogs are constantly barking so others cannot speak to each other," says Theodoridis. ■

INFRASTRUCTURE

Structural biologists share their toys

European network pools resources to unpick the secrets of the cell.

BY EWEN CALLAWAY

To study the components of a cell, from proteins to organelles, scientists need a barrage of high-end equipment that no one laboratory can afford. Researchers typically access the tools they don't have at collaborating labs or national centres. But with the launch of the Instruct network on 23 February, structural biologists across Europe have teamed up to make this sharing more systematic.

Instruct links up 22 structural-biology

centres, allowing researchers to access a range of equipment and expertise in a single request, says Instruct's head David Stuart, a structural biologist at the University of Oxford, UK.

It also presents an opportunity for structural biologists to solicit funding with a unified voice — just in time for Horizon 2020, the European Commission's next research-funding programme, which is gearing up to disburse an anticipated €80 billion (US\$108 billion) in grants between 2014 and 2020.

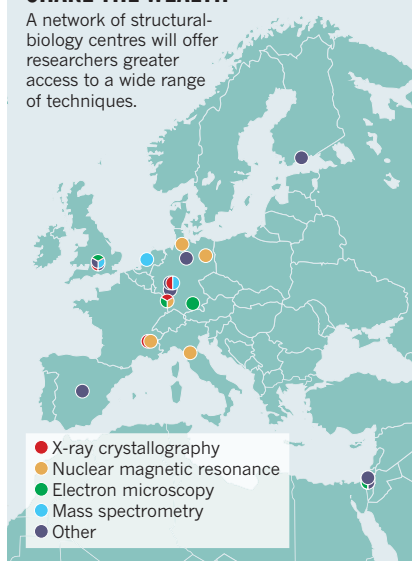
Instruct emerged from Europe-wide

projects that determined the crystal structures of hundreds of proteins and protein complexes. Stuart says that the project will now try to fill in details of the cell at every scale in a way that he likens to zooming towards a location using Google Earth. "The idea is to have a 'Google Cell' approach, so you can drill from the cellular context to the atomic detail," he says.

For instance, a scientist studying herpes infection might start by looking at fluorescent viruses under a light microscope. Switching to electron microscopy would reveal ▶

SHARE THE WEALTH

A network of structural-biology centres will offer researchers greater access to a wide range of techniques.



► the subcellular compartments that the virus exploits. A crystal structure of a key molecular interaction might help to identify infection-blocking drugs.

“Only doing crystallography is not enough. Only doing electron microscopy is not enough. You often need all these technologies, and it’s very hard to have them all in one house,” says Albert Heck, scientific director of the Bijvoet Center for Biomolecular Research at Utrecht University in the Netherlands, which will offer mass spectrometry through Instruct.

Each of Instruct’s facilities will cede up to 20% of their capacity to the programme (see ‘Share the wealth’). Instruct will seek independent peer review of requests for equipment use and will approve access to various facilities, in some cases bypassing the allocation committees at individual centres.

The programme is initially using a subscription model, with eight nations, including Britain, France and Germany, agreeing to pay €50,000 per year to give their scientists access. Currently, the European Commission provides funding only for researchers to travel to these facilities. But Heck hopes that if Instruct is successful, Horizon 2020 will support both access and infrastructure for structural-biology centres, including equipment and staff, removing the need for subscriptions.

Aled Edwards, a structural biologist at the University of Toronto, Canada, is well used to juggling international resources as head of the Structural Genomics Consortium, which has solved more than 1,000 protein structures. He thinks Instruct makes sense. “It gives people access to technologies they don’t have locally, it trains people,” he says. “It’s clearly setting itself up for Horizon 2020 funding. But they deserve it, if they can pull it off.” ■

REPRODUCTIVE BIOLOGY

Egg-making stem cells found in adult ovaries

Discovery could pave the way for new fertility treatments and a longer reproductive life.

BY KENDALL POWELL

It’s time to rewrite the textbooks. For 60 years, everyone from high-school biology teachers to top fertility specialists has been operating under the assumption that women are born with all the eggs they will ever produce, with no way to replenish that supply. But the discovery of human egg-producing stem cells, harvested from the ovaries of six women aged 22 to 33, puts that dogma in doubt.

The work, published online in *Nature Medicine*¹ by Jonathan Tilly and colleagues at Massachusetts General Hospital in Boston, parallels the findings of a Shanghai-based group² that isolated similar stem cells from mice in 2009. However, both this and Tilly’s earlier work in mice³ remained controversial, with many experts sceptical that such stem cells existed.

“This is unequivocal proof that not only was the mouse biology correct, but what we proposed eight years ago was also correct — that there was a human population of stem cells in young adult tissue,” says Tilly.

To address the doubts, Tilly’s team began by developing a more sensitive method for identifying and collecting mouse ovarian stem cells. Their method, based on a technique called fluorescence-activated cell sorting (FACS), attaches a fluorescently labelled antibody to a protein, Ddx4, that is present on the outer surface of the stem cells but not on the surface of the later-stage egg cells or oocytes. The FACS instrument lines up cells in single file and sorts them one by one, separating the labelled ones from the rest; it also gets rid of dead or damaged cells, such as oocytes, in which internal Ddx4 might become accessible to the antibody. This method is more selective than previous isolation methods, which did not get rid of such cells.

Once the team confirmed that it had isolated mouse ovarian stem cells by this method, it set its sights on reproductive-age human ovaries. Yasushi Takai, a former research fellow in Tilly’s lab and now a reproductive biologist at Saitama Medical University in

Japan, supplied frozen whole ovaries removed from sex-reassignment patients, all young women of reproductive age. “It was 9 November when we did the first human FACS sort and I knew immediately that it had worked,” says Tilly. “I cannot even put into words the excitement — and, to some degree, the relief — I felt.”

The cells they pulled out, called oogonial stem cells (OSCs), spontaneously generated apparently normal immature oocytes when cultured in the lab. To look at the development of the putative human OSCs in a more natural environment, the team labelled the cells with green fluorescent protein to make them traceable, and injected them into fragments of adult human ovarian tissue, which were then transplanted under the skin of mice. After one to two weeks of growth, the OSCs had formed green-glowing

cells that looked like oocytes and that also expressed two of the genetic hallmarks of this cell type.

“There’s no confirmation that we have baby-making eggs

yet, but every other indication is that these cells are the real deal — bona fide oocyte precursor cells,” says Tilly. The next step, to test whether the human OSC-derived oocytes can be fertilized and form an early embryo, will require special considerations — namely, private funding to support the work in the United States (federal funding cannot by law be used for any research that will result in the destruction of a human embryo, whatever the source of the embryo) or a licence from the UK Human Fertilisation and Embryology Authority to do the work with collaborators in the United Kingdom.

Evelyn Telfer, a reproductive biologist at the University of Edinburgh, UK, was once sceptical of the mouse work, but has become a believer. “I’ve visited [Tilly’s] lab, seen these cells and how they behave. They’re convincing and impressive,” she says. Telfer, who studies the maturation of human eggs *in vitro*, will work with Tilly to try to grow the OSC-derived eggs to the point at which they are ready for fertilization.

She notes that there’s still no evidence that the OSCs form new eggs naturally in the body.

“I’ve seen these cells and how they behave. They’re convincing and impressive.”



► NATURE.COM

For a video explaining this research, see: go.nature.com/5ub5ma

However, if they could be coaxed in a dish to make eggs that could successfully be used for *in vitro* fertilization (IVF), it would change the face of assisted reproduction.

"That's a huge 'if,'" admits Tilly. But, he continues, it could mean an unlimited supply of eggs for women who have ovarian tissue that still hosts OSCs. This group could include cancer patients who have undergone sterilizing

chemotherapy, women who have gone through premature menopause, or even those experiencing normal ageing. Tilly says that follow-up studies have confirmed that OSCs exist in the ovaries of women well into their 40s.

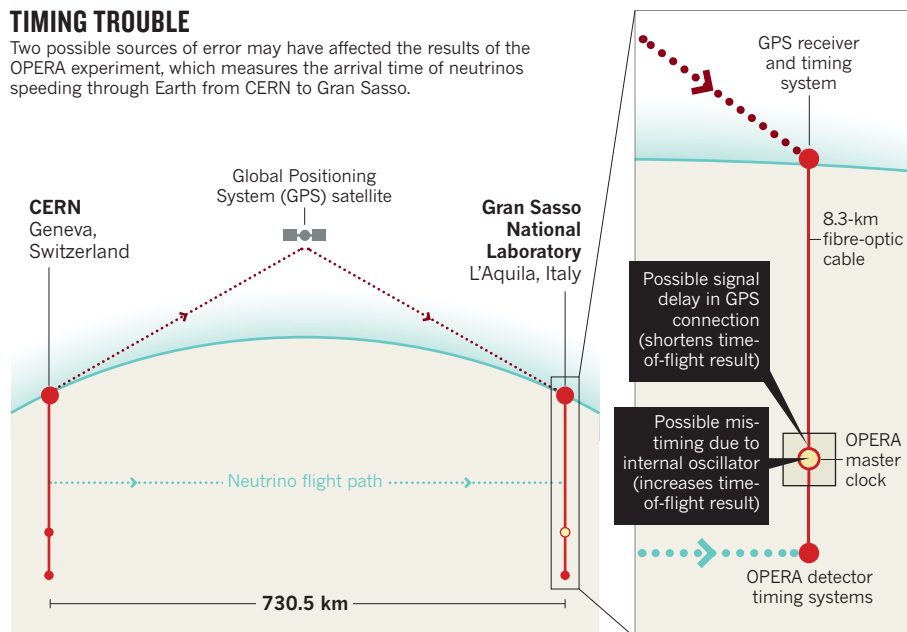
In addition, growing eggs from OSCs in the lab would allow scientists to screen for hormones or drugs that might reinvigorate these cells to keep producing eggs in the body and

slow down women's biological clocks. "Even if you could gain an additional five years of ovarian function, that would cover most women affected by IVF," notes Tilly. ■

1. White, Y. A. R. *et al. Nature Med.* <http://dx.doi.org/10.1038/nm.2669> (2012).
2. Zou, K. *et al. Nature Cell Biol.* **11**, 631–636 (2009).
3. Johnson, J., Canning, J., Kaneko, T., Pru, J. K. & Tilly, J. L. *Nature* **428**, 145–150 (2004).

TIMING TROUBLE

Two possible sources of error may have affected the results of the OPERA experiment, which measures the arrival time of neutrinos speeding through Earth from CERN to Gran Sasso.



experiment. The initial result suggested that the neutrinos were reaching the detector 60 nanoseconds faster than the speed of light would allow. Both potential errors would affect the neutrinos' arrival time, as measured by OPERA's master clock (see 'Timing trouble'). The first is a faulty connection at the point at which the light from a fibre-optic cable brings a synchronizing Global Positioning System (GPS) signal into the master clock. The fault could have delayed the GPS signal, causing the master clock to run slow and thus causing the neutrinos' travel time to appear shorter than it actually was.

"It's a subtle effect," says Autiero, and one that was evident only when the team examined many measurements of signals passing through the connection. Tests of the timing system turned up a second, opposing effect: an oscillator within the master clock that keeps time between the arrivals of synchronization signals was running fast. That would have made the neutrinos' travel time seem longer.

The collaboration says that it has not yet worked out the magnitude of these effects. Autiero says that because of the high profile of the result and the possibility of rumours and leaks, the collaboration wanted to disclose the potential errors promptly. The OPERA team plans to correct the faults and repeat the experiment after CERN's neutrino beam is switched on again in March, following a winter break.

Two independent checks of the measurement are also being considered. One, at Japan's Tokai to Kamioka (T2K) neutrino experiment, would still be valuable despite the doubt cast on the OPERA data, but may now prove harder to fund, says international co-spokesman Chang Kee Jung, a physicist at Stony Brook University in New York. But another, the Main Injector Neutrino Oscillation Search (MINOS) experiment, which fires neutrinos from Fermilab in Batavia, Illinois, to an underground detector in northern Minnesota, will proceed, at a cost of about US\$500,000. "It's never a bad idea to have multiple measurements," says MINOS co-spokesman Rob Plunkett.

Jorge Páramos, a physicist at the Higher Technical Institute in Lisbon, says that the admissions by OPERA point to an honest mistake, albeit one that should have been avoided. "The putative origin of the systematic error reflects the innards of the experiment — something that should have been checked exhaustively before any public announcement," he says. ■

PHYSICS

Timing glitches dog neutrino claim

Team admits to possible errors in faster-than-light finding.

BY EUGENIE SAMUEL REICH

Is it an epic blunder or a textbook demonstration of how science should work? To some physicists, the OPERA (Oscillation Project with Emulsion-tracking Apparatus) collaboration deserves credit for disclosing possible errors in its paradigm-challenging measurement of neutrinos travelling faster than light. "I think we did the right thing to continue to investigate," says Dario Autiero of the Institute of Nuclear Physics of Lyons in France, who presented the original results and notes that the collaboration had spent six months checking its result before its announcement last September.

To others, the revelation shows that the OPERA team went public too soon with its claim that neutrinos from CERN, the European particle-physics laboratory near Geneva in Switzerland, were flouting Albert Einstein's absolute limit on the speed of light as they travelled the 730 kilometres to the OPERA detector at the underground Gran Sasso National Laboratory near L'Aquila, Italy. "I find it embarrassing," says Luca Stanco of the National Institute of Nuclear Physics in Padova, Italy, an OPERA member who initially refused to sign a paper about the result. "Maybe we should have been more cautious and done more checks."

On 23 February, OPERA team members reported two possible sources of error in the



Livestock in the heart of US cattle country will be put at risk if foot-and-mouth disease escapes from the proposed National Bio and Agro-Defense Facility.

BIOSECURITY

Plans stall for biodefence lab

US National Academy of Sciences reassesses risks of high-security work in cattle country.

BY SUSAN YOUNG

For Katharine Bossart, a trip to the lab can involve a 22-hour flight. Bossart, a microbiologist at Boston University in Massachusetts, works on treatments and vaccines for the Nipah and Hendra viruses, which are deadly to both horses and humans. Her research requires the highest level of biological security containment — BSL-4 — but no BSL-4 labs in the United States can accommodate horses, so she collaborates with researchers in Australia.

“If we want to protect large animals from these infections, then we have to test vaccines in them,” says Bossart.

The US Department of Homeland Security (DHS) has broken ground for a facility that would have allowed researchers such as Bossart to work closer to home. The National Bio and Agro-Defense Facility (NBAF) in Manhattan, Kansas, is designed to provide BSL-4 containment for large-animal studies and replace the Plum Island Animal Disease Center, the federal government’s 58-year-old BSL-3 installation off Long Island in New York. But the NBAF’s future has been thrown into question, with no new money allocated for it in the president’s proposed 2013 federal budget and reviewers considering fears about whether it could keep pathogens safely contained in the middle of prime US cattle country.

Last week, the US National Academy of

Sciences (NAS) convened a closed meeting to review a revised risk assessment from the DHS, which it received on 10 February but has not yet made public. Congress ordered the report last year, after the NAS harshly criticized a 2010 assessment, citing factors such as the absence of back-up high-efficiency particulate filters in the building plans; flawed estimates of how quickly an outbreak could be detected and stopped; and poor consideration of the facility’s proximity to metropolitan areas and livestock. Beef producers have been particularly alarmed that the 2010 assessment put the cumulative risk of foot-and-mouth disease escaping from the NBAF over the facility’s projected 50-year lifespan at 70% (see ‘Fear

factor’). The virus that causes the disease spreads quickly and would have a devastating effect on the US cattle industry if it escaped. US research with live foot-and-mouth virus is currently restricted to Plum Island.

Both the latest assessment and the review of it by the NAS, expected by June, are required by Congress before the US\$50 million designated for the NBAF in the 2012 federal budget can be spent.

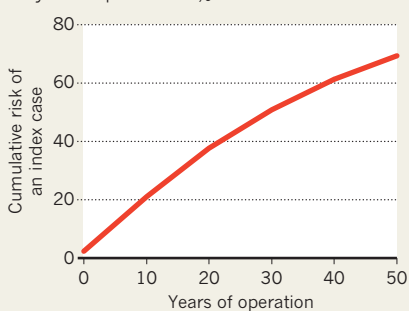
The 2013 budget not only commits no further funds to the NBAF, but also creates a new hurdle, requiring the NAS to examine whether current disease threats justify the facility, which could cost up to \$1 billion to build.

“We understood going into this that issues would arise,” says Ron Trewyn, vice-president for research at Kansas State University (KSU) in Manhattan, which in 2009 helped the state to win its bid to host the NBAF. “Budget is a big deal these days, but given the importance of the NBAF to national security and to protecting our agricultural economy, we are confident that these issues will be worked through and it will move forward.”

Kansas has committed \$105 million in bonds to support the facility, which is touted as an economic boon for the state. The site is adjacent to the KSU Biosecurity Research Institute, a BSL-3 facility that studies animal and plant pathogens. The university is now

FEAR FACTOR

A 2010 risk assessment found that the probability of foot-and-mouth disease escaping from the National Bio- and Agro-Defense Facility during its 50 year lifespan was 70%.



➔ **NATURE.COM**

For more on science and the 2013 US budget, see: go.nature.com/twcvpo

considering whether to send graduate students to Plum Island to build expertise. “We are working with Homeland Security and the US Department of Agriculture on Plum Island to develop the workforce that will ultimately work at the NBAF,” says Stephen Higgs, director of the Biosecurity Research Institute.

But critics of the facility welcomed the decision by President Barack Obama’s administration not to request further funding for it. “We are optimistic,” says Tom Manney, a retired KSU biophysicist who helps to lead a group called No NBAF in Kansas. The group says that a facility that works on highly infectious animal diseases does not belong “in the centre of the food-animal health corridor”.

Opponents voiced their concerns about the project at an NAS public meeting in Manhattan in January. “It is easy for those promoting the facility to argue for the Kansas site because their livelihoods are not at stake,” wrote cattle rancher Paul Irvine in a submitted statement.

What happens next will depend, in part, on the NAS’s judgement of the facility’s risks and benefits. The DHS says that the NBAF is needed to develop countermeasures against bioterrorism — a threat that resonates less now than it did immediately after the anthrax attacks on the United States in 2001. But the department also cites three threats that receive far less public attention. One is the growing likelihood of foreign animal diseases entering the United States as a result of international animal movement from commerce and smuggling. A second is the risk of animal-borne diseases spreading to humans as population growth and dispersal puts people into greater contact with wild animals. The third is the potential for global warming to expand the range of insect-borne diseases.

“Not having a facility like this is almost negligent,” says Higgs. “We have the capabilities to build a facility that will better prepare us in the event of some pathogen coming in. The NBAF will be the shining star in these types of labs.”

Soren Alexandersen, director of a BSL-4 facility in Winnipeg, Canada, that can accommodate small numbers of livestock, says that the challenges of running such labs can be met with technical measures. He adds that although many of the diseases studied in Winnipeg, including the Nipah Virus, are not currently found in North America, preparedness matters. “We have the methods and the facility in place so that we can start working,” he says.

Bossart agrees. “You can’t just flip a switch and make these facilities function,” she says. “If you don’t have these unique capabilities, you are not going to be able to respond if an outbreak occurs.” ■

FUNDING

US disease agency in fiscal peril

Proposed budget changes threaten disease prevention and surveillance programmes.

BY MEREDITH WADMAN

When US President Barack Obama proposed a US\$664-million cut in congressional funding for the US Centers for Disease Control and Prevention (CDC) in his 2013 budget request, he tried to ease the pain by replacing much of it with money from other sources. But only days after the 13 February request, a vote on Capitol Hill made clear just how vulnerable those substitutions are, suggesting that the US public-health agency is on increasingly shaky financial ground.

The proposed cut would come from the part of the agency’s budget that is controlled by Congress and pays for the core operations of the CDC, based in Atlanta, Georgia. These include grants to local, county and state public-health departments to monitor infectious diseases or track food-borne outbreaks. Core funding is also used to maintain the Strategic National Stockpile, a repository of drugs reserved for fighting epidemics and bioterrorism. If Obama’s plan is enacted, the CDC’s congressionally controlled funding will have fallen by roughly 20% since 2010 — a decline that “looks like a disaster waiting to happen,” says Scott Becker, executive director of the Association of Public Health Laboratories in Silver Spring, Maryland.

Cuts to the CDC have already contributed to the loss of nearly 50,000 jobs in state and local health departments since 2008. This year, the administration argues that “efficiencies” will make possible the specific cuts it has proposed in areas such as adult-immunization funding and epidemiological support. But CDC advocates and public-health officials are sceptical. A proposed \$47-million cut to the Strategic National Stockpile “is a lot more than just efficiency. It’s going to cut capability as well,” says Crystal Franco, an associate with the Center for Biosecurity of UPMC in Baltimore, Maryland. “We are reaching the tipping point where preparedness efforts are going to be reversed because of the lack of funding,” she adds.

The CDC’s budget was significantly boosted a decade ago because of concerns over bioterrorism in the wake of the 9/11 attacks. But as a budget vice tightens on the US government, the administration has begun to rely on two other sources to offset cuts to CDC programmes that have historically been funded by Congress (see ‘Cut to the core’).

One source is transfer payments made to the CDC from other public-health agencies. The 2013 budget would boost those payments by \$296 million over this year’s level. But that would require congressional approval, which is far from guaranteed.

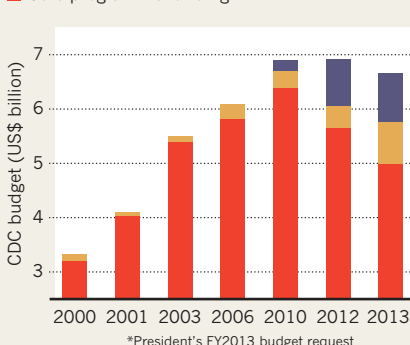
The second source of funds to offset CDC cuts — \$903 million from the Prevention and Public Health Fund (PPHF) — could be even less secure. Established by the 2010 health-reform law, the multibillion-dollar fund is aimed at disease prevention but has become an object of disdain for those opposed to the law. In the current atmosphere of fiscal constraint, the PPHF has become a target for raiding, and already, as part of a payroll-tax-cut extension that was signed into law by Obama on 22 February, Congress has cut the fund by 20%, or \$250 million, in 2013, and by a total of \$6.25 billion to 2025.

The fund “is vulnerable. It could go away quickly,” says James Hughes, a CDC veteran and immediate past president of the Infectious Diseases Society of America in Arlington, Virginia. What’s more, says Hughes, now a professor of public health at Emory University in Atlanta, the fund was intended to launch innovative prevention initiatives, “not to replace core public-health capacity support”. ■

CUT TO THE CORE

The budget of the US Centers for Disease Control and Prevention (CDC) is increasingly floating on funds that may not materialize.

- Prevention and Public Health Fund
- Transfers from other agencies (requires approval by Congress)
- Core programme funding





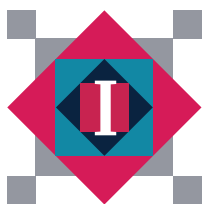
BRIAN DONNELLY STUDIO

Erik Puffenberger, Kevin Strauss and Holmes Morton (from left to right) in front of the Clinic for Special Children, which treats those with inherited disorders.

Genomics, plain and simple

A Pennsylvania clinic working with Amish and Mennonite communities could be a model for personalized medicine.

BY TRISHA GURA



In 2003, Leon and Linda Hoover embarked on a trying medical odyssey by horse and buggy. The Hoovers, of Lewisburg, Pennsylvania, are Mennonites, a religious group that

mostly eschews cars, televisions and other modern conveniences. Along with the Amish communities that dwell in the region, they are the 'Plain people', whose handmade clothes and horse-drawn carriages are an iconic part of the landscape.

But modern medical technology was to become a large part of the Hoovers' lives when they rode to the doctor with their six-week-old

son, Raylon. He had been suffering from a deluge of infections — from croup to thrush to cytomegalovirus. After two months and consultations with various physicians, he received a tentative diagnosis: severe combined immunodeficiency (SCID), which describes a number of genetic disorders that leave children unable to fight infections. Although SCID can be treated by bone-marrow transplantation, refining the diagnosis and finding a matching donor require genetic testing. Physicians shipped blood samples to Duke University in Durham, North Carolina, and sent the Hoovers home to wait. But after three long weeks without results, Raylon's health was worsening, and Leon was getting desperate.

He called the Clinic for Special Children in Strasburg, Pennsylvania, a timber-framed, slate-roofed structure housing an arsenal of modern genetic and genomics instrumentation. Hoover spoke for 45 minutes to Kevin Strauss, a paediatrician there. "You have a SCID baby, and you are at home?" Strauss asked, incredulous. To speed things up, he referred the Hoovers to his colleague Nancy Bunin, now director of stem-cell transplantation at the Children's Hospital of Philadelphia (CHOP) in Pennsylvania. She quickly arranged for genetic testing and a bone-marrow transplant. But it came too late. Raylon died at 6 months and 15 days.

Last September, when the Hoovers had another child, the story was very different. A

test at birth revealed that their daughter, Kendra, had a dangerously low white-blood-cell count and probably had SCID. In tears, Hoover called Strauss on a Sunday at dawn. The paediatrician sprang into action. A midwife collected Kendra's umbilical-cord blood and had it driven the 170 kilometres to his laboratory. Within 12 hours of Kendra's birth, the lab team had detected a single letter change in her *IL7R* gene — the same mutation that had affected her brother.

The next morning, Strauss drove to the Hoover family's sewing-machine shop. There, 16 relatives lined up to give blood in the hope that one could become Kendra's bone-marrow donor. Testing at the clinic revealed that her sister, 11-year-old Ester Mae, was a match, and Kendra received the transplant at CHOP within 16 days of her birth. Kendra's total medical bill was US\$12,000 — compared with \$500,000 for her brother. And she is doing well today.

"This is what genomic medicine is supposed to do," Strauss says. "If you know which people are at risk, you can determine a diagnosis before a child is 24 hours old. You can come up with a treatment based on the genetics."

THE MILLION-DOLLAR INTERPRETATION

The Clinic for Special Children, it seems, has found a way to apply the basic tools of genomics to save lives, money and resources. At the clinic, two paediatricians, a molecular geneticist and their staff have tools such as sequencers, microarrays and a LightScanner, which detects gene mutations. Using these tools, they nimbly stitch together the elements of basic science and clinical practice necessary to move from a blood sample to DNA analysis, all the way to diagnosis and treatment — sometimes in a matter of days.

By doing so, the clinic has achieved what many others in genomic medicine are struggling to do. Although genome sequencing is creeping into clinical care around the world, it has yet to become an integral part of everyday medical practice. "We've talked about the thousand-dollar genome and the million-dollar interpretation," says Eric Topol, a genomicist at the Scripps Research Institute in La Jolla, California. "The challenging bottleneck is the process of trying to nail down which DNA variation is the root cause."

The team at the Clinic for Special Children can negotiate that bottleneck in part because of the unusual community with which it works. The Amish and Mennonites in Lancaster County are descended from a small founder population, have a remarkable knowledge of family history and tend to marry within the groups. This means that they have a high rate of particular genetic disorders, which makes it easier for researchers to trace the causes of such diseases.

But the clinic's success also hinges on its combined clinical and laboratory facilities, and the close relationship that Strauss and

his colleagues have built with the community. The Amish and Mennonites shun technology to varying degrees — some even forbid zips on clothing — but they unanimously support the clinic. If a technology draws people together in "fellowship," says Mark Martin, a Mennonite and member of the clinic board, then the churches will indulge. The community even holds quilt and handicraft auctions to raise funds for the clinic, which last year netted \$500,000 — about one-third of the clinic's budget.

Holmes Morton, who started the clinic in 1989, says that it could serve as a model for personalized medicine in many other small populations with similar genetic histories. "Plain populations are interesting not because they are different," he says. They are interesting because their genetics are the same as those of people everywhere, he says.

Morton, now 61, grew up in Fayetteville, West Virginia. He was a high-school dropout, who spent six years as a boilerman and engineer in the Merchant Marines and the Navy while taking a slew of correspondence courses. He used these to talk his way into Trinity College in Hartford, Connecticut, and from there went to Harvard Medical School in Boston, Massachusetts. The dean of admissions, Morton says, was intrigued by his unusual background.

In 1988, while completing a medical fellowship in metabolic diseases at CHOP, Morton encountered his first Amish patient — a six-year-old boy with a strange form of cerebral palsy. Through a urine test, Morton diagnosed him with glutaric aciduria type 1 (GA-1), which is caused by a defect in a protein-digesting enzyme and can lead to brain damage, severe movement problems and early death unless strict dietary restrictions are observed from

infancy¹. (Morton and Strauss later developed a formula to meet those restrictions.)

Morton knew that there were more Amish children with GA-1 and other treatable genetic conditions who were not receiving care, either because centres were too far away or tests were too expensive. So he decided to open a clinic that could meet the needs of the community. This meant bringing in diagnostic equipment to provide cheap, in-house testing, factoring in physician time for home visits and, most importantly, setting up shop in the heart of the community — within driving distance for a horse and buggy.

BREAKING GROUND

Morton applied for federal funding to cover office space, a computer and a mass spectrometer, but was denied. So he and his wife Caroline decided to take out a second mortgage on their house. Just before they did so, a reporter from the *The Wall Street Journal* wrote about Morton's quest². Within a week, readers had sent in several hundred thousand dollars and Hewlett-Packard had donated equipment. In the grand barn-raising tradition of the Plain people, Amish and Mennonites came together to build the clinic on a plot of land donated by Jake Stoltzfoos, an Amish farmer whose grandchildren Morton had treated. It is probably the only medical centre today with both a hitching post and an Ion Torrent DNA sequencer.

Despite the technology available, Strauss says, one of the clinic's most important tools is "institutional memory". An Old Order Mennonite family came to visit Strauss in 2006, for example. Their six-month-old daughter, Rosalyn, had a cluster of developmental problems. Unable to diagnose her, Strauss, who would eventually take over as medical director of the clinic, wandered downstairs to talk to his lab



Prompt genetic testing and treatment have given Kendra Hoover (front centre) a good prognosis.

director Erik Puffenberger, a molecular geneticist. “I’ve got a girl with a small head and vision problems,” Strauss said. “Does that ring a bell?”

It did. Puffenberger recalled a Mennonite family he had met in the late 1980s when he was working with Victor McKusick, a renowned geneticist and author of *Medical Genetic Studies of the Amish*. The family had six children: all blind, with abnormally small heads. Puffenberger and Strauss tracked down the family to the same cinder-block residence they had occupied all their lives. The mother was in her eighties, still caring for her five surviving children, by then in their fifties and sixties. Strauss collected blood samples and Puffenberger analysed them, using gene-expression analysis and sequencing all the known protein-coding regions of the genome, known as the exome. Puffenberger and his colleagues identified a novel mutation in a gene called *TUBGCP6*, which encodes a protein necessary for proper cell division and seems to explain the condition³.

Although the disease cannot be cured, Rosalyn’s vision problem, which involved the retina detaching from the eye, was treated with surgery to prevent otherwise inevitable blindness. Strauss’s team has since caught the same disorder, before the onset of symptoms, in two of her siblings through newborn screening. One has already received surgery and the other, still an infant, is scheduled to undergo surgery soon.

In addition to those within the Lancaster community, Morton and Strauss examine patients from communities in 27 other states and a handful of other countries. Like any paediatrics practice, the physicians send out some samples for routine testing, but can perform almost any genetic and biochemical test in their own basement laboratory. They now know and can test for the molecular basis of 121 heritable genetic diseases common among the populations they serve. If a patient has none of these — the case for nearly half the children they see — the basic research starts.

Strauss estimates that, in the Lancaster populations, his team will uncover the roots of between 5 and 15 new genetic diseases a year for at least the next decade. Many of those will be present in only a handful of people. About one-half to three-quarters will be treatable, he estimates, especially if detected early, as in the case of Kendra Hoover.

How applicable their approach is for other clinics that are currently adopting genomic tools is an open question, however. Some scientists are sceptical. “It’s a pretty unique situation because they are dealing with a closed

population,” says Leslie Biesecker, chief and senior investigator of the genetic-disease research branch at the National Human Genome Research Institute in Bethesda, Maryland. By ‘closed’, he means that because of its small founder population and intermarriage, Plain people stand to inherit only a relatively small number of rare disorders.

Although conceding that disease genes are easier to identify among the Amish and Mennonites, Strauss argues that the criticism misses the mark. “If isolated populations make it so easy to do this work,” he asserts, “why haven’t other major academic centres nearby that see these patients made the genetic dis-

In the United States, for example, there are pockets of high intermarriage in Appalachia. And Iceland has some 318,000 residents, who are largely descended from a small founder population. Genomics approaches similar to that used at the clinic are already at work for consanguineous populations in the Middle East and Ireland, says Stacey Gabriel, director of cancer and medical sequencing at the Broad Institute in Cambridge, Massachusetts.

“Founder populations are the history of the world,” Morton says. “Whether they are located in a county in West Virginia or a region in Brazil, each population has a genetic make-up that can be studied and served.” That, he

adds, is where the model of the clinic becomes useful.

Moving forward, the team at the clinic hopes to streamline its approach to diagnosing new patients and the roughly 1,000 current patients in the area still without a molecular diagnosis.

For help in proving that the mutations they identify really do lie at the heart of

disease, the team is collaborating with neuroscientist Rob Jinks of the Lancaster-based Franklin & Marshall College. With support from an education grant from the Howard Hughes Medical Institute, Jinks and a group of undergraduate students are working to determine the biological consequences of the mutations the clinic discovers.

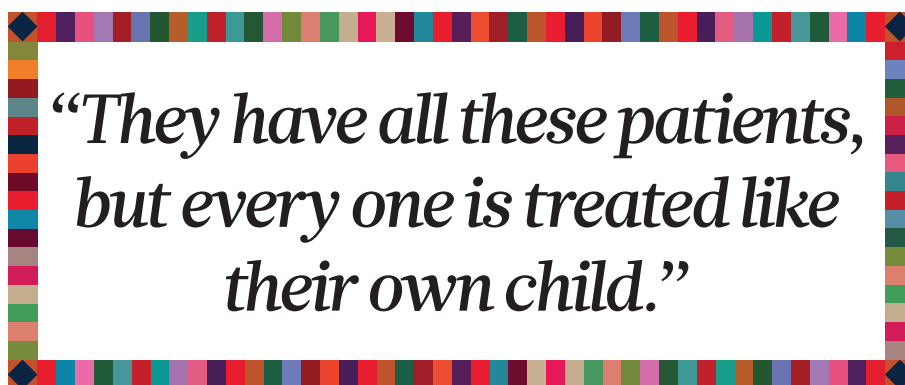
But the clinic team is quick to note that scaling up genomic sleuthing is not the top priority. “We solve the problems for one child, one family at a time,” Strauss says.

The Hoovers have experienced this first hand. Last October, Kendra Hoover lay in her mother’s arms as a needle delivered human immunoglobulin into her scalp. These proteins would help protect Kendra from viral and bacterial infection while the cells transplanted from her sister began to take residence and give her a functioning immune system. Her prognosis is good, and her father, Leon, couldn’t be happier.

“I don’t know if I can put into words, how much they mean to us,” he says. “They have all these patients. So many to take care of. But every one is treated like their own child.” ■ **SEE EDITORIAL P.5**

Trisha Gura is a science journalist in Boston, Massachusetts.

1. Morton, D. H., Bennett, M. J., Seargeant, L. E., Nichter, C. A. & Kelley, R. I. *Am. J. Med. Genet.* **41**, 89–95 (1991).
2. *Wall Street J.* (20 September 1989).
3. Puffenberger, E. G. et al. *PLoS ONE* **7**, e28936 (2012).



coveries that we have?”

The answer, he says, is that other institutions keep the clinic separate from the research, which complicates genomic screening and makes it prohibitively expensive. By contrast, the clinic embeds its laboratory within its walls, certifies it and uses a blanket institutional review board approval from nearby Lancaster General Hospital. That allows Strauss or Morton to order, say, a microarray test to screen for a genetic mutation in a newborn in the same manner that other physicians might order a cholesterol test. The lab’s integration, as much as the community structure, is key to surmounting the problem of million-dollar interpretation.

Puffenberger says that the advantages provided by such populations are shrinking. Many of the team’s recent discoveries were — or could have been — determined by sequencing the exomes of just one patient and his or her immediate family, rather than reaching out to extended families for multiple affected individuals. As other medical centres begin to apply genome and exome sequencing, Strauss, Morton and Puffenberger hope that more will adopt the same sort of integrative style.

Even sceptics such as Biesecker concede that the idea of a general population as a well-mixed melting pot is inaccurate. Instead, the ‘general population’ is actually a mosaic of subpopulations, immigrating together and often living for decades in the same geographical location.

NATURE.COM
Listen to a podcast
about the Clinic for
Special Children at:
go.nature.com/xctunw

GAME ON

Michael Merzenich has a plan for how to convince sceptics of the worth of his brain-training video games: prove that the software can help people with schizophrenia.

BY ERIKA CHECK HAYDEN

The intersection of Mission and Sixth streets in San Francisco's South of Market neighbourhood is considered one of the most crime-riddled in the city. Liquor shops, adult bookshops and single-resident-occupancy hotels inhabit most of the buildings. Homeless people sit on the pavements or shuffle by, many of them showing symptoms of mental illness or drug abuse. Yet behind the walls of an unassuming outpatient psychiatric clinic, researchers are conducting experiments that they believe could fundamentally change the landscape of psychiatric care.

Inside the San Francisco Citywide and Community Focus Center, in a room about the size of a large walk-in wardrobe, two people wearing headphones sit staring at computer screens. Despite the hubbub — the din of a nearby group session, clients milling in the hallway and the internal turmoil caused by their mental disorders — they are mesmerized by the games on the screens. Both have schizophrenia, and they are taking part in a study that aims to determine whether a controversial method of treatment can succeed where modern medicine has largely failed.

The man behind the games is Michael Merzenich, an emeritus professor at the University of California, San Francisco, and a pioneering advocate for neuroplasticity — the

notion that the brain can reshape and remodel its neural pathways even into adulthood. He has gained notoriety for his unabashed promotion of video games designed to tap into that plasticity. These have been marketed to treat everything from reading difficulties in children to driving impairment in elderly people. His zeal has, at times, attracted criticism. There is little solid evidence, say critics, that brain training makes a long-lasting difference in the lives of either ill or healthy people. The language-learning program Fast ForWord, in particular, has rankled some. "It is a waste of money for schools to pay for these programs that, in my reading of the literature, are ineffective," says Martha Farah, a cognitive neuroscientist at the University of Pennsylvania in Philadelphia.

But this year, Merzenich aims to push the envelope further. He and his colleagues plan to start clinical trials that could lead to a first in the brain-training field: approval from the US Food and Drug Administration (FDA).

FDA approval for the software would give people with schizophrenia a new sanctioned treatment option, and grant legitimacy to Merzenich's work. It will also put pressure on the rest of the brain-training industry to prove the value of its products. Making the case will not be easy. But Murali Doraiswamy, a psychiatrist and neurobiologist at Duke University in Durham, North Carolina, says that Merzenich is

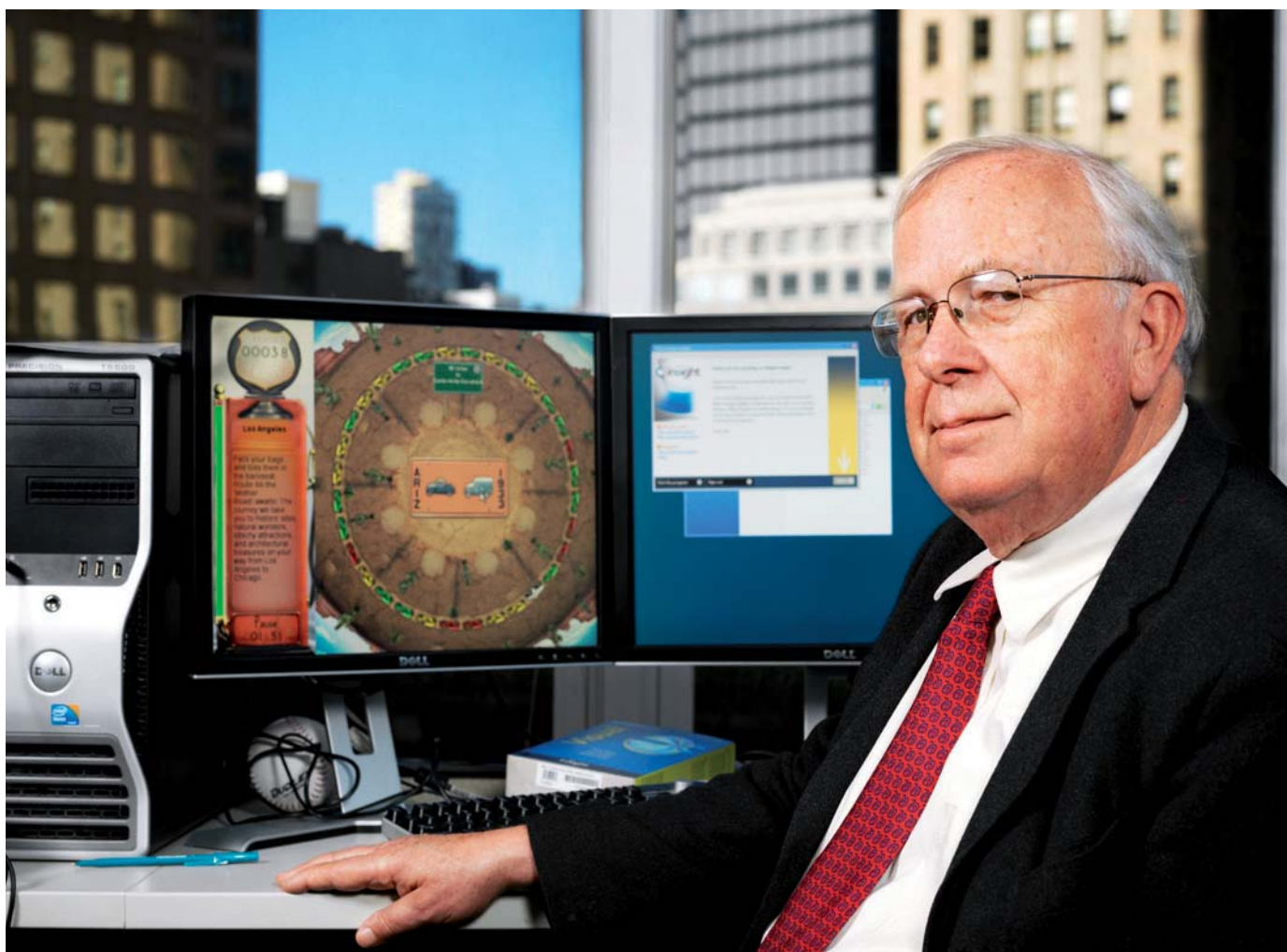
fighting an important battle. "If he can present rigorous results on the outcomes that the FDA deems meaningful, it will help everyone else to focus on achieving those goals and weed out the bad players."

Weeding out the 'bad players' is one of Merzenich's main hopes; he has long felt that the field of cognitive training lacks discipline and is happy to share what he thinks of it: "bullshit-ridden". Even some critics, such as Farah, respect his push for FDA approval. "I think this foray into schizophrenia puts us back in much more familiar territory, where we have systems in place for evaluating whether an intervention is effective."

READY FOR ACTION

A gruff bear of a man, Merzenich grew up as the son of a farmer in rural Oregon. He owns a small farm in Sonoma County, north of San Francisco, where he makes wine and cures olives, and he arrives for a Monday-morning interview with dirt under his fingernails.

Merzenich earned his PhD in 1968 at Johns Hopkins University School of Medicine in Baltimore, Maryland, working with Vernon Mountcastle, who laid the basis for brain-plasticity studies by showing that the cerebral cortex is organized into vertical columns of neurons that respond to particular stimuli. Merzenich went on to demonstrate in monkeys



Michael Merzenich has courted controversy with his brain-training software. Now he is trying to get it approved by the Food and Drug Administration.

CODY PICKENS FOR NATURE

that when neurons that normally process sensory information from one part of the hand lose their inputs, they begin processing input from other parts of the hand¹. Merzenich also mapped the brain systems involved in human hearing — work that eventually led him to co-develop the first cochlear implant, a device that allows people with severe hearing impairments to hear by training their brains to interpret electrical signals sent to the auditory nerve. That device, developed in the 1980s, is now sold by Advanced Bionics in Valencia, California.

His work on the monkeys and the implant showed Merzenich first-hand how the brain could change given the right stimuli, and he began to see rewiring the brain as a way to treat a broad range of medical and societal ills. In 1996, he began selling Fast ForWord through a company he co-founded in Oakland, California, called Scientific Learning. The program aims to train the brain to discriminate sounds better. And that, Merzenich contends, should lead to improved speech processing and reading. In 2003, he formed Posit Science in San Francisco, California, to focus on cognitive training, essentially workouts to sharpen the brain's function.

Posit, like other brain-training firms, has faced scepticism. The brain-training industry — which was projected to grow from US\$265 million in 2008 to between \$1 billion and \$5 billion by 2015 — markets games that

**“WE WILL SEE
BENEFITS.
WE WILL.
BECAUSE WE
ALREADY HAVE.”**

claim to boost skills such as memory or focus in healthy adults. But for those likely to need it most, such as elderly people, there has been no convincing evidence that the games work any better than the mostly free activities that physicians routinely recommend, such as physical exercise, socializing with friends, taking up a new hobby or playing a musical instrument. “Really well-designed clinical trials to test the efficacy of these devices are few and far between. It’s sort of like the Wild West,” says Peter Snyder, a neurologist at Brown University’s Alpert Medical School in Providence, Rhode Island.

Henry Mahncke, Posit’s chief executive, says that the company has shown memory improvements in healthy people aged 60 years and up who use the software^{2,3}. He also criticizes a 2010 paper⁴ in *Nature* that reported a study of people

recruited over the Internet to play a brain-training game. The study showed that benefits didn’t translate into general cognitive improvement. It was widely reported as undermining the concept of brain training, but Mahncke points out that the training was much less intensive than in Posit’s trials and the participants were younger. Still, the criticisms have taken a toll. Posit, for instance, shrank from 86 employees in 2007 to fewer than a dozen last year. The company needs a clear success.

Schizophrenia might seem an odd place for Posit to stake its future because the disorder is complex and can involve symptoms as diverse as hallucinations, delusions, lack of motivation and difficulties in expressing emotions, speech and thought. But the drugs currently used to treat it have frequent and severe side effects. They also don’t alleviate many of the symptoms that can prevent people with schizophrenia from finding work and forming relationships. That means that schizophrenia has a higher social cost than more common psychiatric illnesses, such as depression.

Yet the disease’s intractability is part of what attracted Merzenich; if his treatment succeeds, it could have a big impact. Moreover, researchers are shifting their focus away from the more attention-grabbing symptoms of schizophrenia — the delusions and other signs of loss of contact with reality — and

starting to see the disorder as the result of dysfunctions in basic cognitive processes such as attention, learning, memory, speech and problem-solving.

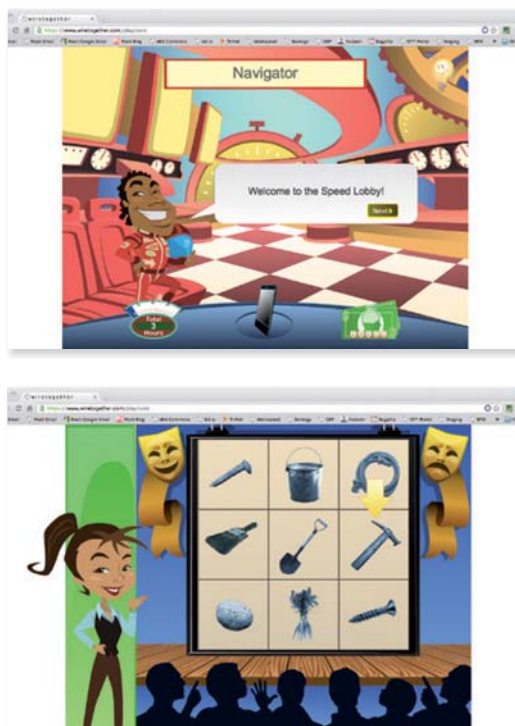
That recognition helped to give rise to a therapeutic approach known as cognitive remediation, which aims to tackle the deficits directly. It also led the US National Institute of Mental Health in Bethesda, Maryland, to commission guidelines for testing interventions in schizophrenia. Called MATRICS (Measurement and Treatment Research to Improve Cognition in Schizophrenia), the guidelines were developed in consultation with the FDA and released in 2005. They spell out which cognitive tests should be used to measure deficits and improvements in seven categories of mental abilities, such as visual learning and problem-solving, and how the tests should be used in clinical trials intended to form the basis of FDA approval. MATRICS was designed to facilitate the development of new drugs, but Merzenich sees it as a road map that could guide trials of brain-training software.

Thanks to the MATRICS guidelines, Merzenich says, “schizophrenia is a relatively simple thing to carry to trial. If you have a significant effect in a clinical population, then you will be given the right to sell the drug.” He is optimistic that his treatment will be approved. “I’ll tell you right now, we will see benefits in all seven MATRICS categories. We will. Because we already have.”

The approach that Merzenich is testing, called Plasticity Assisted Cognitive Remediation (PACR), includes elements from programs that Posit already markets to the general public. One set of exercises focuses on improving the processing of sounds and words through tasks that, for instance, ask the user to discriminate between similar sounds, such as ba and ga, or to decide whether a tone is rising or falling in pitch. Another deploys similar tasks in a visual space — for instance, asking a user to remember where items are hidden, or to spot an object placed in the periphery of his or her visual field. Other exercises, specifically designed for people with schizophrenia, focus on social function.

TRAIN OF THOUGHT

Sophia Vinogradov, a psychiatrist at the University of California, San Francisco, has already run some clinical trials of Merzenich’s software. The goal of treating people with schizophrenia, she says, is to improve fundamental neural processes so that the brain is freed to tackle higher-order cognitive demands such as working memory. By analogy, she says, if you were trying to improve on a lousy tennis serve, you might get better results by focusing on components necessary for the motion, such as balance, coordination and upper-body strength, than by simply repeating your ineffective serve over and over again. People with schizophrenia have



Modules in Merzenich’s software for people with schizophrenia target basic cognitive tasks.

problems at multiple levels of auditory and verbal processing pathways, Vinogradov says. “It makes sense to start with early levels of these pathways and work your way up.”

Some suspect that Posit will have more success in showing benefits for serious disease than for healthy adults. “If you’re very close to maximum performance, you don’t have a lot of room for improvement,” says Doraiswamy. “Whereas for people in, for instance, a pre-Alzheimer’s stage, if you can show 15% to 20% improvement, it might mean postponing the development of Alzheimer’s by six or eight months.”

Vinogradov and others have tested various levels and combinations of these training programs in small groups of people with schizophrenia. She has shown that the most-intensive training tested — 100 hours — can improve global cognition for as long as six months after the training ends⁵. None of the groups showed a significant improvement in symptom severity or quality of life, but Vinogradov speculates that that was probably partly because there were only about ten patients in each group. A small study published last week showed that intensive training can improve people’s performance in a task that assesses the ability to distinguish what is real from what is not. Those who improved most had significantly enhanced social functioning six months later⁶.

When combined with other interventions, such as counselling and job-placement programmes, as the current trial at Citywide clinic is, cognitive remediation approaches seem to help patients function better in society — for instance, by helping them to keep jobs⁷. But

teasing out what role, if any, the software has in these gains is difficult, partly because the programs have been tested in so many different combinations and conditions.

Alice Medalia, a clinical psychiatrist at Columbia University in New York and a pioneer in cognitive remediation, says that she doesn’t think Posit should be seeking FDA approval yet, because it hasn’t shown its software to be any better than other cognitive remediation therapies, and because it is not clear how the programs actually treat schizophrenia. A meta-analysis⁸ of cognitive-remediation studies, including those testing Posit’s software, found no differences in the effectiveness of various programs — or of other methods of non-pharmaceutical remediation.

This calls into question the scientific rationale of the games, Medalia says. “We have no idea what the active ingredient in these computer-based programs is. Saying that neuroplasticity is the active ingredient is not specific.”

Merzenich, however, feels that he has fought long enough to prove the validity of brain training. Now, he says, it is time for regulators to weigh in. Treating schizophrenia with software would mark a change for psychiatry, which tends to focus on dispensing drugs in the first instance. Vinogradov says that the growing realization of drugs’ shortcomings and a shift away from the idea that brain deficits are immutable are sparking desire for alternative options.

“The dominant force in psychiatry has been the focus on treating symptoms, not the underlying dysfunction. The patient is this passive object to whom you give pills, as opposed to actively helping to stimulate constructive interaction with his or her environment,” Vinogradov says.

And Merzenich doesn’t plan to stop with schizophrenia. The Brain Plasticity Institute in San Francisco, another Merzenich-founded company, is studying brain-training software for conditions ranging from Alzheimer’s disease to traumatic brain injury. “If we do this in a disciplined way, with scientific confirmation that is beyond question,” Merzenich says, “we’ll very rapidly evolve into a very important aspect of psychiatric medicine.” ■

Erika Check Hayden is a reporter for *Nature* based in San Francisco, California.

1. Merzenich, M. M. *et al. Neuroscience* **8**, 33–55 (1983).
2. Mahncke, H. W. *et al. Proc. Natl Acad. Sci. USA* **103**, 12523–12528 (2006).
3. Smith, G. E. *et al. J. Am. Geriatr. Soc.* **57**, 594–603 (2009).
4. Owen, A. M. *et al. Nature* **465**, 775–778 (2010).
5. Fisher, M., Holland, C., Subramaniam, K. & Vinogradov, S. *Schizophr. Bull.* **36**, 869–879 (2010).
6. Subramaniam, K. *et al. Neuron* **73**, 842–853 (2012).
7. Wexler, B. E. & Bell, M. D. *Schizophr. Bull.* **31**, 931–941 (2005).
8. Wykes, T., Huddy, V., Cellard, C., McGurk, S. R. & Czobor, P. *Am. J. Psychiatry* **168**, 472–485 (2011).

COMMENT

BIOSAFETY The risk posed by synthetic microbes in the environment **p.29**

ECONOMICS Taxing tobacco, trade and arms could plug the green funding gap **p.30**

HISTORY How John von Neumann drove early computing at Princeton **p.32**

ARCHAEOLOGY Science gives up some of Tutankhamun's secrets **p.34**



D. ORTIV/THE ART ARCHIVE



The hieroglyphic script of the ancient Egyptians mystified scholars until the nineteenth century.

A clash of symbols

Andrew Robinson pieces together the story of who deserves the credit for deciphering the hieroglyphs.

In archaeology, as in other sciences, assigning credit for achievements is often controversial. The decipherment of Egyptian hieroglyphs in the 1820s is an intriguing case. Recognition for this great feat is generally — and properly — given to the French linguist and archaeologist Jean-François Champollion, who is regarded as the founder of Egyptology. But it is widely acknowledged that crucial initial steps were taken by the English polymath Thomas Young.

Having written biographies of both men, I am convinced that the narrowly focused, hot-tempered, impecunious Champollion learnt much from seeds of knowledge planted by the wide-ranging, cool-headed, wealthy Young. Had they chosen to combine forces, rather than to be divided by the defeat of Napoleon Bonaparte (a patron of Champollion), the problem could well have been solved years earlier.

In stark contrast to Champollion's specialization in Egypt, Young was a professional

physician. He is best known as a physicist — for his double-slit experiment showing the interference of light waves and for his modulus of elasticity — and as a physiologist, who theorized about how the eye sees colour. He was also a polyglot, who named the Indo-European language family in 1813, after comparing some 400 languages.

Young became hooked on ancient Egyptian scripts in 1814, when he began to decipher the Rosetta Stone. Since its discovery by Napoleon's army in 1799, the tablet's three equivalent inscriptions — two in unreadable Egyptian scripts (hieroglyphic and demotic), the third in the Greek alphabet — had triggered much activity by scholars. Notable among them were the French Orientalist Silvestre de Sacy — Champollion's teacher — and de Sacy's Swedish student Johan Åkerblad.

What ensnared Young was writing a review of a massive German history of languages, the editor of which asserted that the unknown language of the Rosetta Stone and other

Egyptian artefacts might be written in an alphabet of little more than 30 letters. When he was shown some fragments of inscribed papyri from Egypt, Young could not resist the challenge. In the summer of 1814, he took a copy of the Rosetta Stone inscription to the English seaside town of Worthing to study.

CONJECTURE AND COMBAT

In November of that year, as foreign secretary of Britain's Royal Society, Young saw a letter Champollion had written from Grenoble to the society's president¹. It mentioned Champollion's access to two differing reproductions of the Rosetta Stone, and asked the society to check passages from them against the original in the British Museum in London.

Young replied to the letter, and the following year Champollion read what Young called his "conjectural translation" of the Rosetta Stone, attempted in Worthing and lent to Champollion by de Sacy. Then, the battle of Waterloo and its bitter aftermath intervened. The ►

► royalist de Sacy turned against the republican Champollion, openly advising Young to regard his former student as a potential plagiarist¹. The Young–Champollion correspondence ceased.

In late 1821, by his own admission, Champollion read Young's key work on Egypt, published as a supplement to the *Encyclopaedia Britannica* in 1819. In 1822, the two men met at the academic meeting in Paris at which Champollion famously announced his decipherment of the hieroglyphs, barely mentioning Young. In 1822–23, they warily exchanged letters. But when Young published a book claiming that Champollion's decipherment was an "extension" of his earlier work, a heated argument ensued.

They remained respectful to each other in public. Young was, after all, held in high esteem by French scientists. In 1828, the two even worked together briefly in Paris; Champollion, by now the first curator of Egyptian antiquities at the Louvre, offered to open up his collections and some personal notes on the demotic script to a grateful Young.

In private, Champollion was scathing. Writing from Egypt's Valley of the Kings in 1829, Champollion exulted to his brother in French²: "So the poor Dr Young is incorrigible? Why stir up old matter that is already mummified? ... The Briton can do as he pleases — *it shall be ours*: and all of old England will learn from young France to spell hieroglyphs by a totally different method."

French writers have almost unanimously supported the intellectual independence of Champollion from Young, at least after the decipherment was accepted in the mid-nineteenth century and Champollion became a national hero. Outside France, opinion is still divided.

Some British Egyptologists have suggested that Champollion poached from Young the concept of hieroglyphs as a 'mixed', alphabetic-cum-ideographic, system. The Irish clergyman Edward Hincks, an Egyptologist who later helped to decode Mesopotamian cuneiform, claimed in 1846 that Champollion was guided by Young's alphabetic analysis of the hieroglyphic cartouches for the royal names Ptolemy and Berenice.

Others saw Champollion's subsequent interlocking translations of numerous cartouches as key. A former keeper of Egyptian antiquities at the British Museum, Peter le Page Renouf, argued in 1896 that Young's few readings were too unsystematic to have been crucial. His reaction was echoed in 2005 by Richard Parkinson, current curator of the Rosetta Stone, who wrote³ that "while

Young discovered parts of an alphabet — a key — Champollion unlocked an entire written language".

In my view, the truth lies in between. Champollion was firmly on the wrong track in the first half of 1821. It was Young's 1819



Script decoders: Thomas Young (left) and Jean-François Champollion.

article, especially its tentative hieroglyphic 'alphabet', that reoriented Champollion and set him on the right path. Without it, Champollion might never have made his great breakthrough of 1822.

CRACKING THE CODE

Documentary evidence for Young's influence on Champollion is lacking, but Champollion's inconsistent publication record from 1810–21 is suggestive. In April 1821 (before he read Young's *Encyclopaedia Britannica* article), he published a short study illustrating 700 Egyptian hieroglyphic and hieratic signs. Hieratic is the cursive script derived from hieroglyphic in the third millennium BC. It then gave birth in around 700 BC to demotic, the second Egyptian script on the Rosetta Stone. Champollion's study, *De l'écriture hiératique des anciens Égyptiens*, consisted of a mere seven pages of text and seven plates of illustrations. Despite its brevity, it was inadvertently revealing.

The study stated three conclusions². First, hieratic "is no more than a simple modification of the hieroglyphic system, and differs from it only in the form of its signs". So hieroglyphic was the origin of hieratic, and hence of demotic. Second, hieratic "is in no way alphabetical". Third, hieratic characters "are signs of things and not sounds" — ideographic, not phonetic.

Champollion's first conclusion was right, but had already been published by Young. In an 1815 letter to de Sacy, Young noted a "striking resemblance" between some hieratic and demotic signs and "the corresponding hieroglyphics" in the Rosetta Stone and in some papyrus manuscripts. Champollion had made the same deduction in 1821 on his own, or so he said.

Champollion's second conclusion was

likely to be wrong because it contradicted the work of de Sacy, Åkerblad and Young, who agreed that demotic almost certainly contained alphabetic elements. In the same letter to de Sacy, Young stated that demotic was neither purely ideographic nor a pure alphabet, but a mixture.

Given that error, Champollion's third conclusion, that hieroglyphic and hieratic represented "things" and not sounds, was bound to be incorrect. His published denial of phoneticism in hieratic was a blunder, as well as an admission of how far behind Young he was.

Champollion soon regretted the 1821 publication. He allegedly made strenuous efforts to withdraw all copies. This tale is probably true, given that copies are extremely rare and Champollion presented only the illustrations to Young, who was unaware of the text until later. Most tellingly, Champollion made no reference to it in his 1822 paper announcing the decipherment. The suppression makes sense because Champollion had decided — after reading Young — that there was a phonetic element in the Egyptian scripts after all.

So the decipherment of the Egyptian hieroglyphs required both a polymath and a specialist to crack the code, even if Champollion would never admit it in public. Young's myriad-mindedness provided some key initial insights in 1814–19, but then his versatility slowed his progress.

Champollion's single-mindedness hindered him from arriving at these insights in the same period, but once he got started, his 'tunnel vision' allowed him to perceive the system behind the signs in 1822. Both Young's breadth of interests and Champollion's narrow focus were essential.

Had they collaborated, a 'Champollion–Young decipherment' of the Egyptian hieroglyphs might well have begun in 1815 — and gained acceptance decades before Champollion's hotly contested version triumphed in the 1860s, long after its discoverer's death. ■

Andrew Robinson is a writer based in London, and author of *The Last Man Who Knew Everything: Thomas Young (Oneworld/Pi Press, 2006)*. His biography of Champollion will be published in April 2012. e-mail: andrew.robinson33@virgin.net

1. Young, T. *Miscellaneous Works of the Late Thomas Young, M.D., F.R.S.*, vol. 3 (Thoemmes Press, 2003. Reprint of 1855 edn).
2. Robinson, A. *Cracking the Egyptian Code: The Revolutionary Life of Jean-François Champollion* (Thames & Hudson/Oxford Univ. Press USA, 2012).
3. Parkinson, R. *The Rosetta Stone* (British Museum Press, 2005).

Four steps to avoid a synthetic-biology disaster

Assess the ecological risks of synthetic microbes before they escape the lab, say **Genya V. Dana, Todd Kuiken, David Rejeski and Allison A. Snow.**

Picture vast expanses of land, perhaps in the desert or near the sea, where giant ponds of synthetic microalgae pump out biofuels. Imagine seeds coated with engineered *Escherichia coli* bacteria that induce lateral root growth to reduce soil erosion, the goal of the Project Auxin team in the iGEM competition (see <http://2011.igem.org>).

These may sound like amazing innovations that have no downside, but what if some of the modified organisms persist and spread, despite efforts to control them? Could they disrupt the normal functions of ecosystems by transferring their altered DNA to other microbes? Might they, for instance, increase competition for resources, or disrupt crucial ecological functions?

No one yet understands the risks that synthetic organisms pose to the environment, what kinds of information are needed to support rigorous assessments, or who should collect such data. These questions have been raised before, with genetically modified seeds, for example. But unlike transgenic crops, synthetic microbes will be altered in more sophisticated and fundamental ways (such as elimination of metabolic pathways), making them potentially more difficult to regulate, manage and monitor. They might also have environmental impacts that are difficult to predict.

In July 2011, the Synthetic Biology Project at the Woodrow Wilson International Center for Scholars in Washington DC (including us) assembled a group of synthetic biologists and ecologists to explore the possible risks of introducing novel organisms into the environment, and how to assess these risks (see go.nature.com/pncgxs). These scientists are developing an eco-risk research agenda — the first of its kind for this emerging field — to help move the field forward in a productive fashion, while aiming to avoid serious ecological impacts.

It is urgent that we start this work now. Synthetic biology has already moved out of the lab, propelled by significant public and private investments in organisms modified to produce chemicals, medicines and biofuels. Researchers estimate that the global market for synthetic biology was US\$1.1 billion in 2010. This is expected to increase to \$10.8 billion in 2016, with chemicals and energy constituting the largest share¹. Such a prediction conjures images of a world

of 'living factories' filled with synthetic microbes made up of modular genetic parts. It also raises the possibility that these synthetic microorganisms could escape into the environment, perhaps in large quantities.

RESEARCHING RISK

We propose four areas of risk research that scientists, in coordination with regulatory agencies and other key stakeholders, should address immediately to understand the potential ecological effects of synthetic organisms, and to regulate the use of these organisms properly.

First, differences in the physiology of natural and synthetic organisms will affect how they interact with the surrounding environment. Research in this area could, for instance, look for changes in a synthetic organism's production of toxic substances or other harmful metabolites.

Second, escaped microorganisms have the potential to survive in receiving environments (for years if in a dormant state) and to compete successfully with non-modified counterparts. Researchers will need to consider how such microbes might alter habitats, food webs or biodiversity.

Third, synthetic organisms might evolve and adapt quickly, perhaps filling new ecological niches. It is important to ascertain the rate at which the synthetic organism and its genetic material evolves so as to determine whether the organism could persist, spread or alter its behaviour in natural environments — even those that are seemingly inhospitable.

The fourth area of risk research should focus on gene transfer. Microorganisms are known for their ability to exchange genetic material with other organisms or to take up free DNA from the environment. Understanding this process is important for determining whether synthetic organisms will pass on

properties such as antibiotic resistance, which could pose threats to human health.

Risk-research programmes to cover these four areas will need to provide dedicated, long-term funding to obtain high-quality data, because some experiments can take years — for example, investigating rates of evolution and the persistence of synthetic organisms and their DNA in the environment. Over time, experts can use these data

to develop a model that looks at the potential impacts of synthetic biology from a lifecycle perspective².

Public agencies must link basic and environmental risk research by co-funding projects and requiring grant recipients to work with environmental scientists from the start.

Given the complexity of the research questions, the economic and social value of successful synthetic-biology applications and the potential impact of errors, we think that

a minimal investment of \$20 million to \$30 million over 10 years is appropriate.

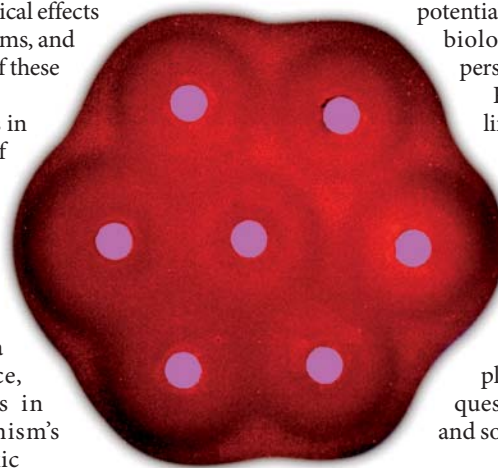
Once released, synthetic organisms cannot be retrieved. It is imperative that funding and research communities take action to prevent future ecological disasters. ■

Genya V. Dana, Todd Kuiken and David Rejeski are at the Woodrow Wilson International Center for Scholars, Washington DC 20004, USA. **Allison A. Snow** is at Ohio State University, Columbus, Ohio 43210, USA.

e-mail: gydana@gmail.com

1. Bergin, J. *Synthetic Biology: Emerging Global Markets* (BCC Research, 2011); available at <http://go.nature.com/wrhta>
2. Davis, J. M. *Comprehensive Environmental Assessment* (US Environmental Protection Agency, 2010); available at <http://go.nature.com/zkardl>

This article was written before G.V.D. was employed by the US Department of State. Views expressed are the author's own and do not necessarily reflect those of the US Department of State or the US government.



The ecological impact of these engineered *E. coli*, in which pink colonies signal to red, is unknown.

NATURE.COM
US implementation of synthetic-biology policies is slow:
go.nature.com/28aqj7

Tax 'societal ills' to save the planet

Funding is a major stumbling block for environmental initiatives, says **Edward Barbier**. Taxing financial transactions or trade in arms, tobacco and fuel might help.

In June, the United Nations Conference on Sustainable Development, 'Rio+20', will aim to green the world economy and, at the same time, eradicate poverty. But progress at Rio+20 is likely to falter over a familiar stumbling block: financing.

Negotiations at conferences such as these are often focused on setting targets for action or making financial commitments, without negotiating the mechanisms by which those funds will be raised. This has resulted in large gaps between intended actions and actual funds, on the order of hundreds of billions of dollars.

Three revenue-raising mechanisms are often proposed: funds, global markets for carbon and ecosystem services, and revenues from global trade. The first two don't seem to hold much promise for generating the hundreds of billions of dollars needed. For example, the Green Climate Fund, agreed at the 2010 Cancun climate talks, promises up to US\$100 billion a year by 2020 to help developing countries to adjust to global warming; yet present international commitments to all climate funding total less than \$2.5 billion. Other funds face similar gaps.

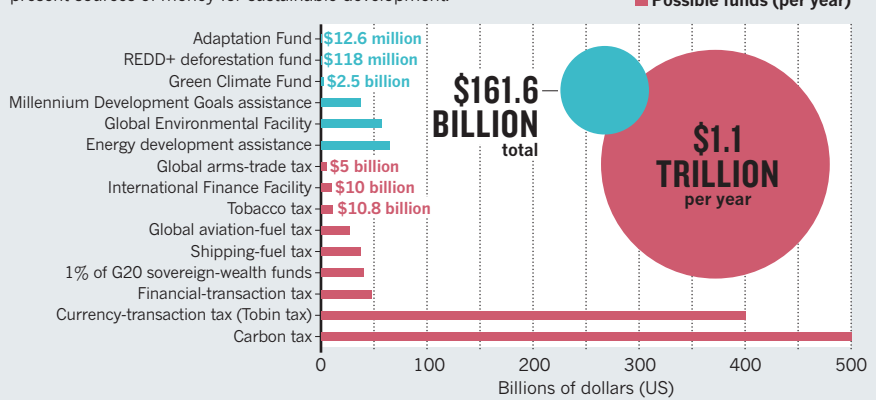
The only global market for ecosystem services yet in operation is the nascent United Nations (UN) programme for reducing emissions from deforestation and forest degradation (REDD+). But the cost of reducing emissions in this way is far greater than the amount that nations could earn through carbon markets to finance the action¹. UN-REDD, a pilot initiative established by several UN agencies in 2008, has raised \$118 million — not nearly enough to make up the shortfall². As for carbon taxes and tradable emissions permits, politics, particularly in the United States, seems to have ruled out these options.

The third option — tapping global trade, including financial transactions — seems far more appealing (see 'Money makers'). There are several ways of doing this. An International Finance Facility (IFF) mobilizes resources from international capital markets by issuing long-term bonds that are repaid by donor countries over 20–30 years. An IFF for immunization, launched in 2006, has so far raised \$3.6 billion. Unlike other global financing mechanisms, an IFF can be started by just a few donor countries, and can be implemented through

NATURE.COM
Edward Barbier on
the Global Green
New Deal:
go.nature.com/czjk5w

MONEY MAKERS

Possible funds from taxation and the financial sector vastly exceed present sources of money for sustainable development.



existing aid institutions, such as the World Bank. However, the amounts raised are likely to be in the tens, rather than hundreds, of billions of dollars.

Alternatively, the UN Conference on Trade and Development has proposed that developing economies invest 1% of their sovereign-wealth funds in regional development banks. Allocating just 1% from the sovereign-wealth funds of G20 countries would raise at least \$40 billion annually³.

CREATIVE TAXATION

Levying taxes on those things that societies find unsalutary, from alcohol to gambling, is a long-established practice. It makes sense to extend the taxation of 'societal ills' to activities that have made a select few so rich and caused so much of the recent financial havoc.

A financial transaction tax (FTT) is one such possible funding source. A small FTT collected on the sale of financial assets, such as stock, bonds or futures, would have a negligible effect on trade, but could raise substantial funds: a tax of 0.1% on equities and 0.02% on bonds could bring in about \$48 billion from G20 member states⁴.

A variant of the FTT is a currency-transaction tax, or Tobin tax, named after James Tobin, the economist who proposed it in the 1970s. Foreign-exchange transactions total around \$800 trillion annually, which means that a Tobin tax of only 0.05% could raise \$400 billion a year⁴.

Other taxes hold lucrative promise. A 10% tax on global arms exports, for example, could raise up to \$5 billion annually⁵. Additional tobacco-sales taxes in G20 and other European Union (EU) countries could generate

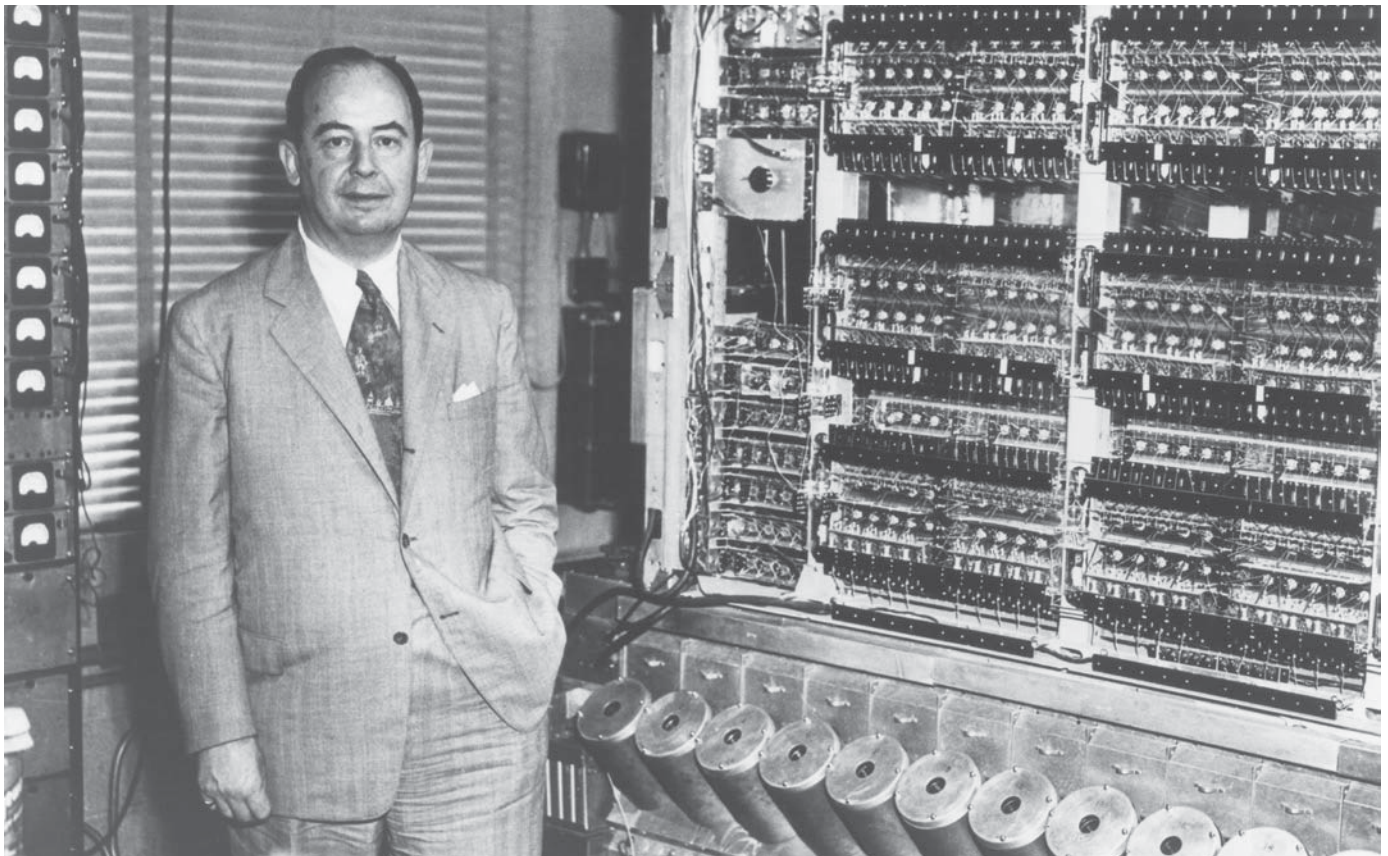
an extra \$10.8 billion; global aviation-fuel taxes \$27 billion; and shipping-fuel taxes \$37 billion³.

The problem with these taxes is that few nations are likely to want to participate: governments will face intense lobbying pressures and the risk of increased black-market activity. Nevertheless, the EU proposed an FTT at the G20 summit in Cannes, France, in November 2011 as a way to raise development funding for poorer countries. Although favourably received by many G20 countries, the proposal failed to secure full backing because of opposition from the United States, the United Kingdom and Canada, all of which worried about the FTT's added burden to their banks. Europe is still pushing for an FTT.

Despite such obstacles, these methods have the best chance of raising the money needed to make a sustainable planet. It is time that such creative funding mechanisms were discussed more seriously at meetings — such as Rio+20 — at which action is proposed. ■

Edward Barbier is in the department of economics and finance at the University of Wyoming, Laramie, Wyoming 82071, USA. His latest book is *Capitalizing on Nature: Ecosystems as Natural Assets*.
e-mail: ebarbier@uwoyo.edu

1. Ebeling, J. & Yasué, M. *Phil. Trans. R. Soc. B* **363**, 1917–1924 (2008).
2. *The Least Developed Countries Report, 2010: Towards a New International Development Architecture for LDCs* (UN Conference on Trade and Development, 2010).
3. Gates, B. *Innovation with Impact: Financing 21st-Century Development* (Gates Notes, 2011).
4. Spahn, P. B. *Broker* **22**, 10–14 (2010).
5. Brzoska, M. *Kyklos* **57**, 149–172 (2004).



PHOTOQUEST/GETTY

John von Neumann led US efforts to develop an early electronic computer, in parallel with similar projects in the United Kingdom.

COMPUTER SCIENCE

Digital dawn

Thomas Misa ponders a history of computing that focuses firmly on John von Neumann and the 'Princeton crowd'.

In *Turing's Cathedral*, George Dyson gives us a personal history of computing's early years. A science historian and son of physicist Freeman Dyson, he recounts a boyhood spent in 1950s Princeton, New Jersey, poking around the forgotten detritus of an early digital computer in the Institute for Advanced Study (IAS). He gives us, too, the engaging tale of how Hungarian mathematician John von Neumann built a computer at the institute. But this is just one of the origin stories of the digital age. Perhaps, at the time, Princeton seemed to be the centre of the Universe.

The IAS was founded in 1930 by Louis Bamberger, a department-store magnate who cashed out just before the Great Depression. Dyson relates how its first director, Abraham Flexner, decided to focus on mathematics and theoretical physics at a historic moment, when pre-eminent Europeans such as Kurt Gödel, Albert Einstein and von Neumann were moving to the United States. Princeton

became a clearing house for Jewish scientists, with IAS mathematicians and physicists housed in Princeton University's Fine Hall for the first years of the IAS's existence.

To this intellectual crucible came Alan Turing, an English graduate student who had solved the 'decision problem' set out in 1928 by David Hilbert, a mathematician at the University of Göttingen in Germany. In 1931, Gödel's incompleteness theorems had ended Hilbert's decades-long quest for a logically complete and consistent arithmetic, and by 1936 Turing and Princeton University mathematician Alonzo Church had independently proved that there were numbers that could be specified but never calculated. Church recognized Turing's paper to be, remarkably,

more general than his own, and invited the young man to study at Princeton for two years. In tackling the problem, Turing had created a hypothetical machine that could compute anything that could be computed. This device, later named the Turing machine, became a central conceptual model for computing. Turing's 1936 paper

On Computable Numbers was consulted so often that Princeton's copy fell apart.

Although Dyson's title anticipates the Turing centennial this June, the book revolves around von Neumann. Dyson says little about their direct interactions; von Neumann admired Turing's work sufficiently to offer him a fellowship at Princeton in 1938, but Turing returned to England. Also left offstage is Turing's wartime work on code-breaking and early computing, which brought him substantial fame beyond mathematics.

Instead, *Turing's Cathedral* relates von Neumann's effort to build an early electronic digital computer. Dyson's account, based on the IAS archives and access to oral histories from von Neumann's family and many close



**Turing's Cathedral:
The Origins of the
Digital Universe**

GEORGE DYSON
Pantheon/Allen
Lane: 2012, 432 pp.
\$29.95/£25



TURING AT 100

A legacy that spans science:
nature.com/turing

colleagues, is compelling. Much of the story is known, but Dyson devotes attention to the phenomenal physical task involved: getting thousands of vacuum tubes to behave and keeping cooling units from icing over, the responsibility of chief engineer Julian Bigelow.

Dyson also emphasizes von Neumann's position in the top-secret US military research community, from the Manhattan Project onwards. He suggests that von Neumann's work on meteorology "offered both a real problem and a perfect cover for the work on bombs". Handwritten computer logs pinpoint midnight runs of weather-forecasting and weapons-designing computer code.

The military funded more than three-quarters of the IAS computer project. Alarmed by such traffic with the real world, IAS scholars relegated the computer engineers to the basement of the mathematics department, next to the men's toilets.

The heyday of the IAS machine was brief: from 1951, when it became operational, to 1954, when von Neumann left for a board position with the US Atomic Energy Commission. He died three years later, aged 53.

The book evokes vivid details of life in Princeton and the personalities of von Neumann and his colleagues, including Willis Ware, Gerald Estrin, Arthur Burks and Herman Goldstine, who between them helped to create much of computer science. But the spotlight on von Neumann will infuriate many students of these complex years.

Dyson overlooks the computing efforts across the Atlantic at the University of Manchester, UK, where a 'stored-program computer' was in operation at least three years before the IAS machine; he also neglects the original programming at the University of Cambridge, UK. Genetic metaphors mislead when he describes a generation of computers as clones of the IAS machine. And Dyson, who covered much of this terrain in his *Darwin Among the Machines* (Basic Books, 1998), over-reaches in asserting that "the entire digital universe ... came into existence in 1951 ... at the end of [Princeton's] Olden Lane".

A balanced account could not ignore William Aspray's *John von Neumann and the Origins of Modern Computing* (MIT Press, 1990), and would fully use Andrew Hodges' insightful biography, *Alan Turing: The Enigma* (Burnett Books, 1983). Turing and Church's early work on logic is documented well in Charles Petzold's compelling *The Annotated Turing* (Wiley, 2008).

Read this book for its memorably told stories of the famous Princeton crowd. But beware its flights of fancy. ■

Thomas J. Misa is a historian and director of the Charles Babbage Institute, University of Minnesota, Minneapolis, Minnesota.
e-mail: tmisa@umn.edu

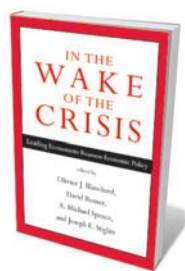
Books in brief



The Righteous Mind: Why Good People are Divided by Politics and Religion

Jonathan Haidt ALLEN LANE 448 pp. £20 (2012)

Morality, above all else, is behind human success, says psychologist Jonathan Haidt. But why do individuals and societies disagree so violently about what morality is? The answer, he says, is that humans are tribal creatures, with minds designed for "groupish righteousness". Haidt discusses how morals arise from intuitions, not reason, and how this leads to both nobility and hypocrisy. He argues that right-wing politicians have a built-in advantage when they appeal to those intuitions by invoking liberty, loyalty and sanctity.



In the Wake of the Crisis: Leading Economists Reassess Economic Policy

Edited by Olivier J. Blanchard, David Romer, A. Michael Spence and Joseph E. Stiglitz MIT PRESS 174 pp. £13.95 (2012)

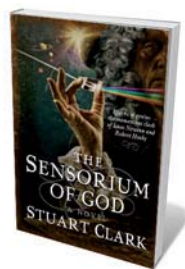
Why did the global economy melt down in 2008? And how can it be fixed? Here, 23 economists, including three Nobel prizewinners, seek the causes of the crash and the feeble recovery, and suggest policy responses. They focus on macroeconomic issues such as the control of inflation and financial-sector regulation. The hunt for an answer will take years, says Olivier Blanchard, chief economist at the International Monetary Fund, but it is clear that we are in a "brave new world".



To Forgive Design: Understanding Failure

Henry Petroski BELKNAP PRESS 408 pp. \$27.95 (2012)

When a plane crashes or a bridge collapses, faulty engineering is the usual suspect. But in seeking the roots of failure, we should look beyond design, says engineer Henry Petroski. We must probe the political and economic imperatives that shape purposes and use. In this follow-up to his influential *To Engineer is Human* (Vintage, 1985), Petroski argues that accidents such as the Deepwater Horizon oil spill are the result of faults as much in "human machinery" as in mechanical devices. He praises software developers for learning from structural engineering about how to report and analyse mishaps.



The Sensorium of God: A Novel

Stuart Clark POLYGON 272 pp. £12.99 (2012)

In 1679, astronomer Edmond Halley seeks out Isaac Newton to help him decipher planetary motion. But the great physicist has less lofty preoccupations: natural philosopher Robert Hooke is demanding recognition for ideas that Newton has passed off as his own. Stuart Clark's novel, the second in a projected trilogy about the giants of physics and astronomy — the first, *The Sky's Dark Labyrinth* (Polygon, 2011), starred Tycho Brahe and Johannes Kepler — sets collaborations and disputes in Britain's Glorious Revolution of 1688. The supporting cast includes John Locke and Robert Boyle.



American Genesis: The Evolution Controversies from Scopes to Creation Science

Jeffrey P. Moran OXFORD UNIVERSITY PRESS 216 pp. \$29.95 (2012)

When historian Jeffrey Moran began teaching in Kansas in 1998, he thought the 1925 Scopes 'monkey trial', in which a Tennessee teacher was prosecuted for teaching evolution, was a curiosity. Yet in 1999, his state's education board removed evolution from its curriculum. Moran discusses how US anti-evolutionism has persisted. Attitudes towards evolution touch on every point of social friction, he says, including gender, race and the North-South divide.



ARCHAEOLOGY

Golden boy

Jo Marchant uncovers a mixed hoard in a history of Tutankhamun and the discovery of his tomb.

Archaeological finds don't come much more impressive than the contents of Tutankhamun's tomb. When Howard Carter glimpsed those "wonderful things" in 1922, his discovery captivated the world — inspiring everything from jewellery and music to a proposal to name an extension to the London Underground Tootancamden, because it would pass through Tooting and Camden Town. Tutmania resurfaced in the 1970s, when a record-breaking touring exhibition of the burial goods radically changed the museum business.

Joyce Tyldesley's *Tutankhamen* rides a more recent wave of interest, including two major scientific studies of Tutankhamun's mummy, a new museum tour and a string of high-profile documentaries. Tyldesley, an Egyptologist, has written not a biography of the boy king, but a broad history of the discovery.

Tyldesley divides her book into two sections. The first, and meatiest, is concerned with what we know of the king himself — including the research on his remains. The second section addresses the cultural construct that has sprung up around him since the discovery, from archaeological conspiracies to the infamous curse.

Today, Tutankhamun is ancient Egypt's best-known ruler, his golden funerary mask recognized around the world. But in his lifetime, Tyldesley tells us, he was a minor king who came to the throne as a child and ruled for a decade at most, dying at around the age of 18. The only pharaoh's tomb ever discovered nearly intact (although it was robbed twice in antiquity), his resting place yielded a breathtaking 5,398 objects, from golden thrones and chariots to a lock of his grandmother's hair.

But Tutankhamun's fame rests on more than that otherworldly cornucopia. He belongs to the romanticized Amarna period, which began when his predecessor,

Akhenaten, threw out Egypt's traditional religion to worship only Aten, the Sun disc. And Tutankhamun's youthful demise set the scene for endless speculation and conspiracy theories.

Tyldesley's account of Carter's momentous find and the clearance of the tomb is brought to life with contemporary quotes and colourful details.

She tells how, shortly after the discovery, Carter bought almost 6 kilometres of cotton wool and several thousand light bulbs in Luxor, then converted burial chambers near Tutankhamun's into a photographic darkroom, laboratory and "luncheon tomb".

The story continues with the unwrapping and autopsy of Tutankhamun's mummy, and the scientific analyses done on it since. The findings of the original autopsy are described — from the king's penis (5 centimetres long, bound in the erect position) to a large scab on his left cheek. But Tyldesley's treatment of more recent analyses is disappointing. Scans in 2005 suggested that the boy king had a broken leg, and DNA tests in 2010 provided evidence that he was inbred, and died of malaria. These studies are the subject of intense debate among experts and are the only new scientific material in the book, yet they are skipped over in a few cursory paragraphs.

Instead, there is a lengthy discussion of the Egyptological evidence for Tutankhamun's family relationships, with tomb paintings and graffiti suggesting an exhausting eight candidates for the pharaoh's mother. The rather unsatisfying conclusion seems to be that we can't say much more about Tutankhamun's life than Carter could in the 1920s.

We still don't know who his parents were, although skull shapes and blood groups suggest that he was closely related to a mummy found in a nearby tomb — who may or may not be Akhenaten. Nor do we know how he died. Tyldesley includes a brief suggested biography anyway, plumping for the fairly safe choices of Akhenaten and a prominent



Tutankhamen:
The Search for an
Egyptian King
JOYCE TYLDESLEY
Basic Books/Profile:
2012. 336 pp.
\$29.99/£18.99

"Tutankhamun's youthful demise set the scene for endless speculation and conspiracy theories."

Tutankhamun's burial mask has become an icon.

1996-98 ACCUSOFT/ROBERT HARDING WORLD IMAGERY/CORBIS

wife called Kiya as Tutankhamun's parents.

After hundreds of books on this subject, the promise on the cover that Tyldesley is "shedding new light" on the boy king was always going to be hard to keep. But she is such a gifted storyteller that perhaps it doesn't matter. Her writing is crystal-clear and charmingly irreverent — she describes Akhenaten's revolutionary reign as "17 years of royal navelgazing" — and she shares intimate anecdotes, such as how Carter and his patron, Lord Carnarvon, broke into the burial chamber ahead of the official opening, covering up their entry point with a basket lid. She puts what little we know about Tutankhamun into context, giving a fascinating discussion of the discovery's social history.

So how does Tyldesley think Tutankhamun died? She dismisses murder, and speculates that the teenager died in a hunting accident. She assigns particular significance to one of his most iconic possessions: the remains of an "ostrich-feather fan" inscribed with hunting scenes. Perhaps the king met his end in pursuit of these notoriously dangerous birds.

It is a nice idea, except that the fan may have nothing to do with ostriches. Tyldesley doesn't mention that in 1978, two US biologists who attended the touring exhibition realized that several animal-related objects from the tomb had been misdescribed by Egyptologists. From the feathers and hunting scenes on the fan, they concluded that Tutankhamun was actually chasing the smaller, and less deadly, Arabian bustard.

And there is the problem — and perhaps the eternal fascination — with this enigmatic king. The harder we look for answers, the deeper the mystery gets. ■ [SEE COMMENT P.27](#)

Jo Marchant is a freelance writer based in London. Her forthcoming book is on Tutankhamun's mummy.
e-mail: jomarchant26@yahoo.co.uk



Hunting scenes on the king's fan offer clues to his death.

THEATRE

Darkness made visible

Daniel Cressey reflects on a play that uses astronomy and medicine to probe what it means to see.

UK theatre company Sound&Fury takes a lab-like approach to its art. Fascinated by the effects of sensory immersion on audiences, it explores mental and physical frontiers — confined spaces in its nuclear-submarine drama *Kursk*, or the effects of anaesthetic on consciousness in *Ether Frolics*. Total or partial darkness has featured in several of the company's productions, including its latest offering, the compelling one-man play *Going Dark*.

Sound&Fury's previous dark dramas have taken audiences from the battlefields of ancient Troy (*War Music*) to a whalers' lifeboat drifting around an unforgiving Pacific (*The Watery Part of the World*). In *Going Dark* — which completes a UK tour this month at London's Young Vic theatre — blindness, hallucinations and astronomy are used to illuminating effect.

Audience members find themselves peering into a planetarium. Under a projection of the Milky Way and often in darkness, this is an exploration of sight through vastly different lenses: the human eye, and the giant telescope arrays through which astronomers access the Universe.

The planetarium's science communicator, Max (played by John Mackay), is diagnosed early on with retinitis pigmentosa, an eye condition that can lead to blindness and hallucinations. He gradually becomes unable to see his own son — or the constellations, real or recreated, that are the focus of his life.

There is a scientific edge to the play. Hattie Naylor's script bristles with astronomical facts, and Dominic ffytche of the King's College London Institute of Psychiatry advised on Max's hallucinations. Co-director Tom Espiner developed a lay grasp of astronomy as a resident artist in the physics and astronomy departments at the University of Birmingham, UK, in 2007.

Max's creeping blindness is partly a commentary on a long-standing nemesis of stargazers, says Espiner: "We suffer from a kind of cultural myopia created by the 24-hour light pollution that hides the stars."

It is not just an inability to see, but the very nature of how we see that Sound&Fury delves into here. "The relationship between eye and brain is a kind of model for the way we



In *Going Dark*, John Mackay is a planetarium employee who loses his sight.

see the Universe," says Espiner. "Seeing is indirect, processed and interpreted by the brain before sight happens. Astronomical observation relies on huge instruments processing spectra of light far outside our perception into information we can process."

There are more teasers in the play's mining of metaphor. Max's bizarre hallucinations range from geometric shapes to human figures: the result of Charles Bonnet syndrome, a side effect of his medical condition. The audience experiences a taste of this when Max, looking into a mirror, sees his face horribly distorted — a vision that the audience shares, thanks to a tiny projector embedded in the set.

After the claustrophobia of *Kursk*, *Going Dark* moves the struggle to inner space — an environment no less extreme. All told, this is a play that asks profound questions about humanity's anchor points in reality, says Espiner.

"What we thought we knew about the Universe changed after Galileo, and changed again with Hubble: the shape, size, levels of reality all shifted," he says. "As he loses his sight, Max has to negotiate a new universe: and we give the audience a sense of what that might be like." ■

Daniel Cressey is a reporter at Nature.

Correspondence

Informed consent: cultural differences

Efforts to improve informed-consent forms for clinical trials should not overlook difficulties faced by patients in non-Western settings (*Nature* **482**, 16; 2012).

Institutions in developed countries are expanding clinical trials in Africa and Asia, and most focus on the signing of the consent form rather than on the exchange of information between researchers and potential participants.

Information should be culturally adjusted, taking local factors into account. These might include degrees of illiteracy, native dialects of ethnic minorities, a lack of suitable vocabulary, a preference for communal decision-making, and stigmatization by local authorities if people do not sign.

Alternative ways of obtaining consent could be used more often, including witnessing oral consent, verifying the quality of information for participants, or using videos to present the study and record consent.

Revising consent forms could help to reassess the concept of globalized bioethics, preventing controversial practices such as those used in researching the genetics of the Havasupai tribe (J. Couzin-Frankel *Science* **328**, 558; 2010).

Gilles Guerrier, Didier Sicard, Paul T. Brey *Institut Pasteur du Laos, Vientiane, Laos.*
guerriergilles@gmail.com

Informed consent: meet patients' needs

We agree that informed-consent forms should be improved (*Nature* **482**, 16; 2012) beyond shortening and simplifying them. Participants need to know whether a trial is addressing an unmet clinical need, and that the test is necessary to answer a key question.

For example, a 'me-too' drug that seeks marketing

authorization by comparison against a placebo, rather than against alternative products, will not solve a patient's unmet need. Neither is it in participants' best interests for a trial to establish the acceptability of a drug with inferior efficacy that provides no other advantage. If the informed-consent form were to explain these points, it is likely that far fewer patients would agree to participate.

Patients also need to understand why randomization is the only scientific and ethical way to provide meaningful results when a trial's outcome is unpredictable. Any other strategy, such as giving or withholding an untested drug, would be arbitrary and would not clarify the uncertainty.

Details of the therapeutic rationale, and of the pharmacological and toxicological properties of the experimental intervention, should be provided to trial participants on request.

Informed consent should reassure patients that trials aim to help them, and will allow onto the market only products that do so, rather than products that just meet commercial expectations. There is a long way to go in cleaning up clinical trials before the trust of patients is regained.

Silvio Garattini, Vittorio Bertele, Rita Banzi *Mario Negri Institute for Pharmacological Research, Milan, Italy.*
vittorio.bertele@marionegri.it

Citations: not all measures are equal

The scientific community needs to be aware of the limitations of Google Scholar's personalized citation reports. Clicking on 'My citations' on the site may offer a nice ego boost, but I would not recommend using the reports for decisions that could affect careers.

Google Scholar overestimates the number of citable articles (in comparison with formal citation services such as Scopus and

Thomson Reuters) because of the automated way it collects data, including 'grey' literature such as theses. For my own publications, for example, Google Scholar yields 38% more citations and boosts the *h*-index by 26%.

A citation report for one of my articles revealed that Google Scholar had counted as independent citations four web pages on which authors had posted copies of their articles, plus one listing only an article title; and one to a paper in which my name didn't appear. Personalized searches by my colleagues exposed comparable errors.

These drawbacks might also allow unscrupulous individuals to use such tactics to inflate their citation reports, particularly as independent vetting is blocked by password access.

Scott L. Hooper *Ohio University, Athens, USA.* hooper@ohio.edu

Citations: results differ by database

Databases such as Thomson Reuters' ISI Web of Science, Scopus, Google Scholar and Microsoft's Academic Search allow authors to compute their own citation statistics, but they yield inconsistent results.

The discrepancies come from differences in information sources and in temporal citation coverage. Web of Science and Scopus, for example, provide citation data only for their indexed journals, giving different coverage for the number of journals, precursor articles and fields of academic research — often with regional biases (such as European versus US sources). Google Scholar includes all journals (indexed, free access and popular science), conference proceedings, books, theses, reports, local press and electronic sources — all subject to variable degrees of control and scrutiny.

A debate is crucial on how these tracking tools compare and should be used, given that their indiscriminate usage has

potentially negative implications for academic careers.

Carles Alcaraz, Sofia Morais *Institute of Agrifood Research and Technology (IRTA), Sant Carles de la Ràpita, Spain.*
carles.alcaraz@gmail.com

Track illegal trade in wildlife

Illegal wildlife trade is the second-largest black market worldwide, after narcotics. More effective strategies must be devised to intercept the first links of the wildlife-trade chain and beyond.

A lack of resources in tropical countries often undermines existing legal frameworks for preventing wildlife trading. Local governments pay scant attention to the trade because it is not perceived as a major threat to biodiversity or to human well-being. A large volume of wildlife trade is international — increasing the risk of biological invasions and of spreading zoonotic diseases.

National environmental agencies should collaborate to centralize the collection and organization of local data to feed into international wildlife-trade databases such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). This would help to identify the species that are most threatened by trade, determine major harvesting sites and routes, and locate sources of demand and supply — as well as revealing the extent of enforcement and of currently unreported shipments. Governments from developing and developed nations could then weigh in with policy improvements.

Luís Felipe Toledo *State University of Campinas, São Paulo, Brazil.*
toledolf2@yahoo.com

Marianne V. Asmussen, Jon Paul Rodríguez *Venezuelan Institute for Scientific Research (IVIC), Venezuela; and EcoHealth Alliance, New York, USA.*

FORUM Planetary science

In search of biosignatures

An analysis of the intensity and polarization of sunlight reflected by Earth reveals signatures of life on our planet. What prospects are there for using similar measurements to find life on planets outside the Solar System? Planetary scientists offer some answers. [SEE LETTER P.64](#)

THE PAPER IN BRIEF

- Ground-based spectropolarimetry can provide better knowledge of the atmospheres and surfaces of planets than can conventional spectroscopy.
- Sterzik *et al.*¹ (page 64) apply the method to Earthshine — sunlight reflected by Earth as seen by the Moon,

and then reflected back to Earth.

- The authors' analysis reveals the contribution to Earthshine of clouds, oceans and even vegetation on Earth's surface.
- The technique might therefore now be used to detect vegetation and other signatures of life on extrasolar planets.

A long way off

CHRISTOPH U. KELLER

Sterzik and colleagues¹ used the highly sensitive 8.2-metre-diameter Very Large Telescope in Chile to detect life on Earth. This may sound like killing a mosquito with a cannon, but it is really a breakthrough: it is the first time that terrestrial life has been detected using a method that can, in principle, also be used to find life on planets outside our Solar System. But we are still far from detecting life on exoplanets remotely.

By applying spectropolarimetry to Earthshine, Sterzik *et al.* were able to determine the

fraction of clouds, oceans and even vegetation on the sunlit part of Earth as seen from the Moon. Spectropolarimetry is a technique that provides not only the intensity spectrum of a light source (spectroscopy), but also the polarization state of the light (polarimetry). In this case, the authors measured the linear polarization content of Earthshine — that is, the degree to which its electric field vibrates in a fixed plane. By comparing the observed signal with models of the polarization signal that had different fractions of clouds, oceans and vegetation, the authors could select the model that best matched the observations.

Linear-polarization measurements of starlight reflected by exoplanets will soon become a common tool for characterizing other worlds.

Both the European SPHERE instrument at the Very Large Telescope² and the US Gemini Planet Imager³ will provide linear-polarization observations that can extract polarized exoplanet light from the much brighter, but unpolarized, starlight. Although these instruments will detect only large gas planets similar to Jupiter and Saturn (Fig. 1), the polarization signal will contain vital information about their atmospheres.

The European Extremely Large Telescope — which, in about 10 years' time, will collect 23 times as much light as the Very Large Telescope — will be able to detect rocky planets possessing liquid water and atmospheres that may allow life to exist. But this huge telescope will probably be unable to obtain the spectropolarimetric measurements of rocky planets that Sterzik and colleagues have obtained for Earth. This is not because the telescope is too small, but because Earth's atmosphere makes it difficult to see a faint planet close to an extremely bright star. To obtain these measurements, a sophisticated space telescope such as the planned New Worlds Mission⁴ is needed.

In the meantime, we must also measure the degree of circular polarization of light from planets. Light is circularly polarized when its electric field describes a corkscrew motion about its direction of propagation. It is well known that molecules produced by living organisms show only one form of handedness (homochirality). When unpolarized light is reflected by biological materials, homochirality imprints a small amount of circular polarization on the reflected light. For example, chlorophyll produces a particular spectrum in the degree of circular polarization⁵. Circularly polarized spectra are therefore an even better tool than linearly polarized spectra for remotely detecting extraterrestrial life — assuming that it also favours homochirality.

Sterzik and colleagues¹ used a large, ground-based telescope to obtain indirect polarization spectra of Earth on two days. Such observations cannot be obtained continuously, as these large telescopes are heavily oversubscribed. Successful Earthshine observations also require specific positions of the Moon and Sun. It is therefore difficult to obtain much

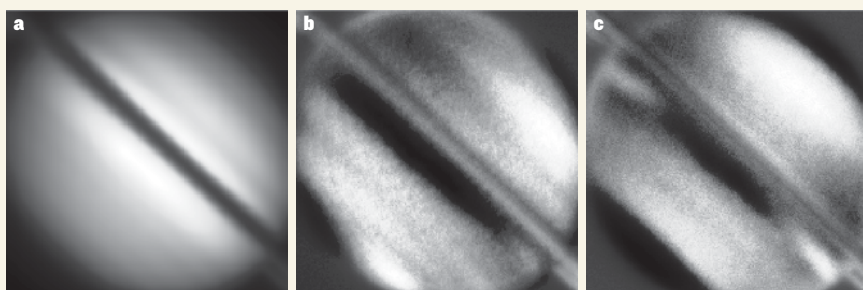


Figure 1 | Saturn's polarization. **a**, Conventional image of Saturn. The planets' rings are visible as a dark band because the rings were almost edge-on at the time of the observations. **b**, **c**, Linear-polarization images of the planet for two polarization orientations. Black and white indicate negative and positive polarization signals; grey corresponds to no polarization signal. Two features stand out in the polarization measurements: the polar regions and a band under the rings. The polar regions are known to have different atmospheric properties from the rest of the planet, but the band under the rings has so far eluded explanation. Sterzik and colleagues' similar analysis¹ of linearly polarized light from Earth revealed signatures of life on our planet. (Images were obtained with the Extreme Polarimeter at the 4.2-metre-diameter William Herschel Telescope on La Palma, Canary Islands.)

better measurements from the ground than those achieved by the authors. However, all of these problems could be avoided by building a small telescope on the Moon that directly and accurately monitors Earth's linear and circular polarization spectrum at all times. By learning how to extract biosignatures, including those of seasons and different types of vegetation, from such measurements, we would be much better prepared to understand future polarization measurements of exoplanets that may harbour life.

Christoph U. Keller is at the *Leiden Observatory, Leiden University, NL-2300 RA Leiden, the Netherlands.*
e-mail: keller@strw.leidenuniv.nl

Benchmark data

DAHPNE M. STAM

Almost 40 years ago, using ground-based polarimetry and a numerical code, researchers⁶ found that Venus's bright cloud layer consists of tiny droplets of sulphuric acid. This result convincingly demonstrated the strengths of polarimetry of reflected sunlight for characterizing planetary atmospheres. Sterzik and colleagues¹ now show that spectropolarimetry of Earthshine provides a benchmark for studying Earth-like exoplanets.

Today, numerical codes similar to the one used to study Venus's atmosphere⁶ calculate polarization spectra of various types of exoplanet. These spectra are used to optimize the design and observational strategy of instruments on future telescopes with a view to studying exoplanets. To ensure that the codes are predicting realistic spectra, given a model planetary atmosphere, they must be tested against real measurements of planets that have well-known atmospheric compositions and structures, such as Solar System planets.

But validating exoplanet codes against polarization measurements of Solar System planets is far from trivial, because the polarization signal of a planet depends strongly on how the Sun illuminates the planet. From Earth, we can observe the outer planets — Mars through to Neptune — only at times when their Earth-facing hemisphere is almost fully illuminated by the Sun (as with a full Moon). And at this geometry, their polarization signal is, unfortunately, almost zero. Only Venus and Mercury (and the Moon) can be observed from Earth at gibbous (more than half but less than full), quarter or crescent phases, when their polarization signal can be strong. Because exoplanets orbit other stars, they can be observed at phases that are favourable for polarimetry.

For exoplanets with Earth-like atmospheres and surfaces, codes can be tested against

polarization spectra of Earth's sky, as measured from the ground, and against polarization measurements from the Earth-observing satellite instrument POLDER. Such local observations cannot, however, capture the rich variation in aerosol and cloud particles, cloud coverage and types of surface that make up a whole-Earth observation, which would be representative of an Earth-like exoplanet observation. Might features such as polarized rainbows, which should be typical signs of liquid-water clouds⁷, and the polarized glint⁸ of sunlight on a water surface, be extracted from a planet's spectropolarimetric signal? Or will the mixture of oceans, continents and clouds dilute them beyond detection?

Sterzik and colleagues' spectropolarimetric measurements of Earthshine¹ offer an opportunity to test numerical codes for Earth-like exoplanets. The authors compare their measurements against calculated spectra⁹. Although the agreement in spectral features is good, there is a significant difference between the slopes of the observed and calculated spectra, especially at the redder wavelengths, where the planetary surface and clouds contribute more strongly to the polarization. The difference in slope is not unexpected, because the available calculated spectra⁹ represent a limited sample of surface types (vegetation and ocean) and include only one cloud type. In addition, the patchiness of the surface and clouds is not taken into account in the numerical code.

Improved spectral calculations should at least include the polarized reflection by the Sahara Desert and the scattering of sunlight

by ice clouds. Indeed, the differences between the measured and best-fit calculated spectra emphasize the sensitivity of spectropolarimetry to atmospheric and surface properties, and hence the value of spectropolarimetry for (exo)planet characterization.

Future spectropolarimetry of Earthshine can reveal whether Earth's oceans flash glints when skies are clear. To confirm the presence of rainbows in Earth's polarization signal, Earthshine should be measured across a narrow range of phase angles as seen from the Moon, a difficult feat. Such measurements could be obtained by a spectropolarimeter located on the Moon that would catch the Earthshine directly. In this way, the detrimental and rather unknown contribution of the lunar reflection to the Earthshine signal would also be eliminated. ■

Daphne M. Stam is at *SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, the Netherlands.*
e-mail: d.m.stam@srn.nl

1. Sterzik, M. F., Bagnulo, S. & Palle, E. *Nature* **483**, 64–66 (2012).
2. www.mpia.de/SPHERE
3. <http://planetimager.org>
4. <http://newworlds.colorado.edu>
5. Sparks, W. B. *et al. J. Quant. Spectrosc. Radiat. Transfer* **110**, 1771–1779 (2009).
6. Hansen, J. E. & Hovenier, J. W. *J. Atmos. Sci.* **31**, 1137–1160 (1974).
7. Karalidi, T., Stam, D. M. & Hovenier, J. W. *Astron. Astrophys.* **530**, A69 (2011).
8. Williams, D. M. & Gaidos, E. *Icarus* **195**, 927–937 (2008).
9. Stam, D. M. *Astron. Astrophys.* **482**, 989–1007 (2008).

SOCIAL SCIENCE

Carrot or stick?

Rewards and punishments can cajole people into cooperating, but they are costly to implement. A theoretical study finds that, when participation in group activities is optional, punishing uncooperative behaviour is the cheaper method.

SIMON GÄCHTER

The philosopher John Locke once wrote¹, “Good and evil, reward and punishment, are the only motives to a rational creature”. Although Locke was referring primarily to the discipline of children, reward and punishment are motivational forces for behaviour across many domains of social life. Understanding the consequences of such ‘carrots and sticks’ is a core topic in the behavioural sciences, particularly in studies of cooperation^{2–8} — behaviour that benefits others or the group at a cost to the cooperating individual. Many problems in modern human societies, from interactions in the workplace to tackling climate change, require genetically unrelated individuals to cooperate in situations

in which collective welfare is jeopardized by individual self-interest. So how do rewards and punishments curb selfishness and help to maintain social order? A paper by Sasaki *et al.*⁹ in *Proceedings of the National Academy of Sciences* helps to answer these questions.

Theoretical and experimental research on the evolution of cooperation has concentrated on punishment, with relatively few studies investigating reward^{7,10–12}. Furthermore, most studies have focused on ‘peer punishment’, in which defectors are punished by group members^{3,7,8,10–14}. However, the findings of peer-punishment studies may not be broadly applicable to modern human societies, which have developed formal sanctioning systems, whereby rewards and punishments are carried out by rule-bound institutions rather than by

individuals (Fig. 1). Sasaki *et al.* address this with a theoretical assessment of how institutionalized reward and punishment systems regulate cooperative behaviour, and at what relative cost.

The authors used evolutionary game theory for their analysis — a theoretical framework in which strategic behaviours can be analysed in the context of evolutionary selection pressure. Their analysis used the ‘public goods game’, in which people can either cooperate or defect, and in which cooperation is collectively beneficial, but defection is better for self-interest. People learn by observing others and they emulate successful individuals, so strategies that yield greater pay-offs proliferate. The model in this study compares one institution that rewards cooperators and another that punishes defectors. The authors studied reward and punishment under two conditions: in one, all individuals were forced to take part; in the other, participation was voluntary.

Sasaki and colleagues found that during compulsory participation the two incentives (punishment or reward) lead to the same outcomes if they are very small or very large. It seems that if either of the incentives is too small, cooperation cannot be achieved because a population of cooperators can be invaded by defectors. If they are large enough, both types of incentive can lead to a population of cooperators. However, differences arise when the incentives are of only intermediate value. Punishments of intermediate severity produce stable populations of either defectors or cooperators, whereas rewards of intermediate value lead to stable mixed populations in which only partial cooperation is achieved.

The authors then changed the rules of the game to allow individuals to opt out, which causes the outcomes to change remarkably. The game now considers three strategies: non-participation, defection and cooperation. In this game, very small incentives lead to an unstable pattern of non-participation with bursts of cooperation. When the incentives are very large, a stable uniform population of cooperators emerges — as it did during forced participation. The most remarkable outcome of the study occurs when intermediate incentives are offered and participation is voluntary. In this situation, slightly increasing the severity of the punishment above a very low level results in stable populations of cooperators. By contrast, rewards of at least medium size are needed to cause a shift in the population from a majority who opt out of participation to a stable mixture of cooperators and defectors. When participation is voluntary, only



Figure 1 | Punishment comes cheap. In modern societies, rule-bound institutions, such as the legal and education systems, use punishments and rewards to encourage cooperative behaviour. Sasaki and colleagues' theoretical analysis⁹ shows that, when individuals can choose whether or not to participate, punishment is the less costly method.

very large rewards could generate a stable and uniform population of cooperators.

This noteworthy theoretical finding has practical implications. Sasaki and colleagues' model can be used to calculate the size of the incentive needed to achieve a stable level of cooperation, although such calculations will be difficult to apply in reality. However, the authors' analysis does raise several issues that are relevant for understanding modern human societies. First, they show that voluntary participation crucially influences the relative cost of reward and punishment. But in modern societies, people can hardly opt out of the law. This might suggest that the results of games played with enforced participation are more applicable. Second, real-life institutions do not work perfectly; for example, punishment or reward may not be correctly implemented. It remains unclear how these imperfectly applied incentives might affect cooperation. Third, law enforcement in reality typically relies on punishment rather than reward. This study may have identified a reason why punishment has become the default in societies — because punishment is a cheaper and more reliable way of inducing cooperation than is reward.

Other studies of human behaviour have also found that cooperation is strongly influenced by changes in the size of the incentive¹³. In experimental studies, reward and punishment induce similar levels of cooperation when the incentive is very large^{11,12}. The threat of a strong punishment can achieve cooperation at a very low cost¹⁴. For an intermediate level of incentive, punishment can induce greater cooperation than reward¹², but not consistently so¹⁰. Finally, cooperation breaks down rapidly if both forms of incentive are removed⁸.

Although these studies were conducted in

settings of forced participation and peer punishment (or reward), their findings are encouragingly similar to Sasaki and colleagues' theoretical analysis⁹. These similarities suggest that experimental analyses conducted in a manner closer to the framework presented by Sasaki *et al.* — institutional delivery of incentives and assessment of voluntary participation — might further enhance our understanding of how reward and punishment maintain social order.

The political scientist Mancur Olson recognized the importance of reward and punishment for cooperation when he argued in his seminal 1965 study¹⁵ that “the recalcitrant individual can be ostracized, and the cooperative individual can be invited into the center of the charmed circle.” But Olson, Locke and other early scholars who wrote about reward and punishment in human social

affairs relied on casual observation and introspection. Collectively, Sasaki and colleagues' study, the related experimental analyses and the potential investigations that arise from them are an example of how researchers today are much better equipped to combine rigorous theoretical and experimental analyses to understand sticks and carrots¹⁶. And this is a rewarding situation indeed. ■

Simon Gächter is in the Centre for Decision Research and Experimental Economics, University of Nottingham, Nottingham NG7 2RD, UK.
e-mail: simon.gachter@nottingham.ac.uk

- Locke, J. *Some Thoughts Concerning Education* (Cambridge Univ. Press, 1889).
- Ostrom, E., Walker, J. M. & Gardner, R. *Am. Polit. Sci. Rev.* **86**, 404–417 (1992).
- Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
- Sigmund, K. *Trends Ecol. Evol.* **22**, 593–600 (2007).
- Ule, A., Schram, A., Riedl, A. & Cason, T. N. *Science* **326**, 1701–1704 (2009).
- Bowles, S. & Gintis, H. *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton Univ. Press, 2011).
- Balliet, D., Mulder, L. B. & Van Lange, P. A. *Psychol. Bull.* **137**, 594–615 (2011).
- Chaudhuri, A. *Exp. Econ.* **14**, 47–83 (2011).
- Sasaki, T., Brännström, Å., Dieckmann, U. & Sigmund, K. *Proc. Natl Acad. Sci. USA* **109**, 1165–1169 (2012).
- Sefton, M., Shupp, R. & Walker, J. M. *Econ. Inquiry* **45**, 671–690 (2007).
- Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. *Science* **325**, 1272–1275 (2009).
- Sutter, M., Haigher, S. & Kocher, M. G. *Rev. Econ. Stud.* **77**, 1540–1566 (2010).
- Nikiforakis, N. & Normann, H.-T. *Exp. Econ.* **11**, 358–369 (2008).
- Gächter, S., Renner, E. & Sefton, M. *Science* **322**, 1510 (2008).
- Olson, M. *The Logic of Collective Action* (Harvard Univ. Press, 1965).
- Poteete, A. R., Janssen, M. A. & Ostrom, E. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice* (Princeton Univ. Press, 2010).

PALAEOBOTANY

In the shade of the oldest forest

The uncovering of a large soil surface preserved under sediment for 390 million years has exposed plant remains which show that the world's earliest forests were much more complex than previously thought. [SEE LETTER P.78](#)

BRIGITTE MEYER-BERTHAUD
& ANNE-LAURE DECOMBEIX

Think for a minute about 'early' life on land. Complexity is probably not the first thought that springs to mind. Botanists also tended to consider the earliest forests to be simple entities composed of a single type of tree. But on page 78 of this issue, Stein *et al.*¹ challenge this view with the description of uncovered plant remains from the ancient forest of Gilboa in New York state. They identify three distinct plant types (or components) in this Middle Devonian ecosystem, which dates to around 390 million years ago, and thereby re-evaluate the ecology of this early environment.

The Gilboa forest is iconic. The discovery² of this "oldest fossil forest" was made in the 1920s at the Riverside Quarry site in Gilboa and other nearby sites, where spectacular sandstone casts formed by the fossilization of stumps of *Eospermatopteris* — the earliest known trees — were found. Stein *et al.*³ had previously discovered specimens representing the trunk and crown of these trees, and assigned the *Eospermatopteris* genus to the Cladoxylopsida, an extinct class of plants related to ferns. The trees had a slender trunk six metres or more in height that bore a crown of leafless and short-lived branches. These were continuously renewed as the plant grew. *Eospermatopteris* trees were assumed to be the only trees of the Gilboa forest.

But when access to the forest's fossils increased after the removal of backfill from the Riverside Quarry site in 2010, researchers discovered a 1,200-square-metre palaeosol surface — one that had been preserved under sediment. Because most of the plant fossils in this surface were found still located where they had grown, Stein *et al.*¹ were able to define plant types and reconstruct structural features of the forest.

The authors identified three large-plant components with apparent differences in growth habit, abundance and phylogenetic relationships (Fig. 1). These components were the previously identified cladoxylopid tree *Eospermatopteris*; a large

rhizomatous plant (one with underground stems growing horizontally⁴) belonging to the extinct aneurophytalean progymnosperms; and a tree with bark similar to that of the lycopsid trees that inhabited coal swamps about 310 million years ago, but for which the authors found too few and fragmentary remains for proper reconstruction.

The discovery that the Gilboa forest did not consist solely of cladoxylopid trees is remarkable. Palaeoecological studies of other Devonian-period sites describe early vegetated

terrestrial landscapes partitioned into a 'two-dimensional' suite of patches growing side by side, each composed of closely related plants with similar morphologies and life traits, and adapted to the same environmental conditions^{5,6}. This structure of Devonian landscapes has almost become a dogma in palaeobotany, but Stein and colleagues' report¹ provides the first direct evidence that some early forests contained widely divergent groups of plants. Furthermore, the differences in the growth patterns of the plants adds a third spatial dimension to the forest ecosystem structure: below the Gilboa trees was an understory of progymnosperm plants with underground, horizontally growing stems and aerial axes.

The combination of a large exposed surface area and *in situ* plant remains at the Riverside Quarry provided sufficient fossil evidence to identify the vertical structures of the Gilboa forest. Such conditions are so rarely met that Stein and colleagues' findings¹ have the power to cast doubt on the supposed simple organization of other early terrestrial ecosystems.

The authors' report also provides evidence to support an alternative view of the growth habit of aneurophytalean progymnosperm plants. These plants were assumed to be bushy and to have a shallow and limited root system⁷. However, subsequent calculations showed⁸ that the shoots of *Tetraxylopteris* plants — the progymnosperm genus probably represented at Riverside Quarry — were not biomechanically capable of the self-support seen in the shoots of bushes and trees. Stein and colleagues' new observations support this result. In addition, they show that the aneurophytalean subterranean system — consisting mainly of rhizomes up to 15 cm in diameter — comprised a large amount of wood and had significantly more mass than previously estimated.

This finding is noteworthy for at least two reasons. First, the *Eospermatopteris* trees — the most conspicuous above-ground members of the Gilboa forest — had a limited amount of wood (if any)⁹; therefore the description of the woody rhizomes adds credibility to the hypothesis that, in early land plants, wood did not evolve as an adaptation for mechanical support¹⁰. Second, the progymnosperm architecture challenges widely accepted theories that there was a straightforward correlation between a gradual increase in plant stature above ground and a gradual increase in size of the corresponding underground parts during the Devonian period⁷.

In addition to their findings on plant types, Stein *et al.*¹ studied sediment deposits around the forest, which existed near the shoreline of an inland sea. Previous studies had likened the Gilboa forest environment to a tranquil swamp, but



Figure 1 | Plant complexity in ancient forests. Plant remains of the Gilboa forest date from the Middle Devonian period, approximately 390 million years ago. Stein *et al.*¹ identify three large components of the forest that are affiliated with separate groups of plants. The two best-preserved components are the *Eospermatopteris* trees (taller aboveground trees), which are related to the ferns, and the rhizomatous aneurophytalean progymnosperms, which grew underground and sprouted aboveground. The morphology of the base and the crown of plants of the third component (depicted as the smaller trees), which seem to be lycopsid-like trees, are unknown.

the observations of Stein *et al.* suggest that the forest was periodically affected by brutal episodes of sea-level rise, which killed some of its plant life. Such disturbances represent major constraints for ecosystems, and can result in evolutionary selection for specific life strategies in plant species. Notably, plants that have adapted to disturbed habitats are generally small in size¹¹. But if Stein and colleagues' observations of a disturbed environment are correct, the occurrence of large trees in the Gilboa forest would go against this trend, and represent an especially intriguing case of plant evolution. ■

Brigitte Meyer-Berthaud and Anne-Laure Decombeix are at the *Unité Mixte de Recherche 'Botanique et bioinformatique de l'architecture des plantes', CNRS-CIRAD, 34398 Montpellier, France.*
e-mails: meyerberthaud@cirad.fr;
anne-laure.decombeix@cirad.fr

1. Stein, W. E., Berry, C. M., Hernick, L. V. & Mannolini, F. *Nature* **483**, 78–81 (2012).
2. Goldring, W. *NY State Mus. Bull.* **251**, 50–72 (1924).
3. Stein, W. E., Mannolini, F., Hernick, L. V., Landing, E. & Berry, C. M. *Nature* **446**, 904–907 (2007).
4. Beentje, H. *The Kew Plant Glossary* (Royal Botanical Gardens Kew, 2010).

5. Hotton, C. L., Hueber, F. M., Griffing, D. H. & Bridge, J. S. in *Plants Invade the Land* (eds Gensel, P. & Edwards, D.) 179–212 (Columbia Univ. Press, 2001).
6. Cressler, W. L. III in *Wetlands Through Time* (eds Greb, S. & DiMichele, W.) 79–102 (Geol. Soc. Am. Spec. Pap., 2006).
7. Algeo, T. J. & Scheckler, S. E. *Phil. Trans. R. Soc. Lond. B* **353**, 113–130 (1998).
8. Speck, T. & Rowe, N. P. *Phil. Trans. R. Soc. Lond. B* **358**, 1473–1485 (2003).
9. Meyer-Berthaud, B., Soria, A. & Decombeix A.-L. in *The Terrestrialization Process* (eds Vecoli, M., Clement, G. & Meyer-Berthaud, B.) 59–70 (Geol. Soc. Lond., 2010).
10. Gerrienne, P. *et al. Science* **333**, 837 (2011).
11. Grime, J. P. *Am. Nat.* **111**, 1169–1194 (1977).

CATALYSIS

Disguise gets a reaction

Every organic molecule has a sea of carbon–hydrogen bonds, so fishing out just one of these bonds for a reaction is difficult. Using a common chemical group as bait provides a solution to the problem. [SEE LETTER P.70](#)

DANIELLE M. SCHULTZ & JOHN P. WOLFE

The beneficial properties of many medicines, materials and consumer products are greatly influenced by the locations and types of chemical groups within those molecules — especially 'functional' groups, such as hydroxyl or amino groups. Typically, the introduction of a desired functional group into a molecule is achieved through interconversion reactions, by which one existing reactive group is exchanged for another. However, this approach can involve many steps, making it costly and time consuming.

On page 70 of this issue, Simmons and

Hartwig¹ report that some functional-group interconversions can be sidestepped by using an iridium catalyst to convert certain carbon–hydrogen (C–H) bonds in molecules directly to carbon–hydroxyl (C–OH) bonds. Their finding provides an extremely useful way to convert alcohols (which contain one hydroxyl group) to 1,3-diols (which contain two hydroxyl groups separated by three carbon atoms). 1,3-Diols are important chemical motifs found in polymers and pharmaceuticals, as well as in the 'skeleton' of many structurally complex, naturally occurring compounds.

The reactions that Simmons and Hartwig report are examples of transition-metal-catalysed C–H bond functionalization², a

synthetic strategy in which otherwise unreactive hydrogen atoms in C–H bonds are replaced directly with functional groups. Such reactions are powerful tools for improving the efficiency with which complex molecules can be synthesized. In practice, however, these reactions are challenging to realize: because organic molecules generally contain many different C–H bonds³, achieving a reaction at just one bond amid all the others can be tricky. Furthermore, each C–H bond in a molecule has a different level of reactivity — although catalysts can be used to functionalize more-reactive C–H bonds in preference to less-reactive ones⁴, it is difficult to selectively functionalize the less-reactive bonds.

One way around this problem is to use a functional group in a molecule as a directing group — a handle to which a metal catalyst can bind and then guide a reaction to a specific C–H bond⁵. The drawback with this approach is that the specialized directing groups required are often not present in the target product of a synthetic route. Additional synthetic steps are therefore needed to introduce and remove the directing group.

Because of this limitation, there has been considerable interest in using functional groups that are commonly found in organic molecules, such as hydroxyls, amides or carboxylic acids, as directing groups for C–H bond functionalization reactions^{6,7}. This strategy has proved successful for reactions of the relatively reactive C–H bonds on benzene rings^{7,8}. But the use of these common functional groups to direct the selective functionalization of less-reactive C–H bonds in saturated hydrocarbon chains (alkyl chains) is considerably more challenging, and has been achieved only in simple molecules⁶.

Enter Simmons and Hartwig¹, who have developed a transformation in which hydroxyl groups on alkyl chains are used to direct a highly selective iridium-catalysed C–H bond functionalization reaction (Fig. 1). This is a notable achievement because hydroxyl groups bind only weakly to transition metals (such as iridium), and have a tendency to take part in unwanted side reactions in the presence of many metal catalysts.

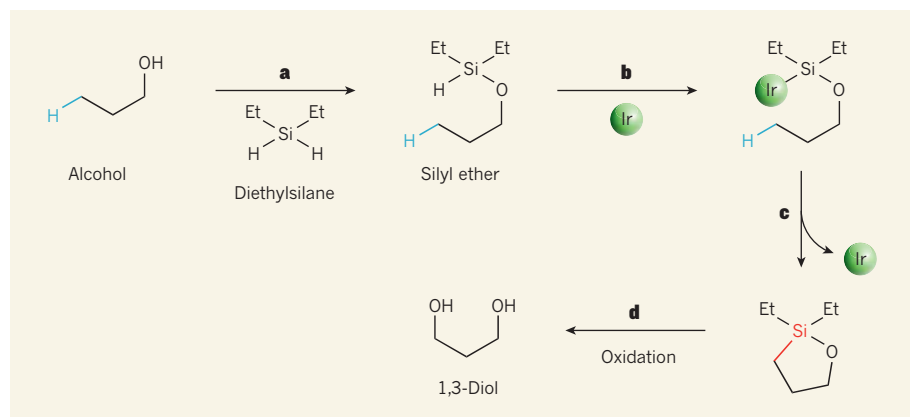


Figure 1 | An iridium-catalysed C–H functionalization reaction. Simmons and Hartwig¹ report a reaction that converts alcohols into valuable 1,3-diols by replacing a specific C–H bond with a C–OH bond. **a**, The starting alcohol is first converted into a silyl ether by reaction with diethylsilane. Et, ethyl. **b**, **c**, The silyl ether forms a strong attachment to an iridium catalyst (green sphere), and directs the replacement of a specific C–H bond (blue) with a carbon–silicon bond (red). The catalyst is released as the carbon–silicon bond forms. **d**, A subsequent oxidation step, performed in the same flask as steps **a**–**c** without purification of the reaction product from those steps, yields the desired 1,3-diol.

The key to the approach was to ‘disguise’ a reactant’s hydroxyl group by converting it into a silyl ether — a silicon-containing group that is known to bind to metal catalysts (Fig. 1). The silyl ether serves as a directing group by forming a strong attachment to the iridium catalyst and directing the formation of a carbon–silicon bond at a specific C–H bond elsewhere in the reactant. In a subsequent step carried out in the same reaction vessel, Simmons and Hartwig removed the silyl ether mask by oxidizing the carbon–silicon bond to a C–OH bond, liberating the desired 1,3-diol product.

The authors demonstrated the power and utility of their method by converting a variety of simple and complex alcohols to 1,3-diols. In every case, the functionalization occurred at a single C–H site three carbon atoms away from the directing group, even if more-reactive C–H bonds were present elsewhere in the molecule. Impressively, the site selectivity remained high even when Simmons and Hartwig subjected structurally complex natural products to the reactions. They were therefore able to convert readily available natural products bearing a hydroxyl group into other natural products that are more difficult to obtain, and to generate new analogues of natural products.

A particularly striking example is the authors’ synthesis of methyl hederagenate (see Fig. 3b of the paper¹) — a precursor of the natural product hederagenin, which has anti-inflammatory, antifungal and antitumour properties⁹. Although the starting material for the synthesis (commercially available methyl oleanate) contains 49 C–H bonds, only one of these bonds was selectively functionalized by their three-step reaction. Previously, the most efficient synthesis¹⁰ of hederagenin required ten steps from a starting material closely related to methyl oleanate.

Although Simmons and Hartwig’s method is very useful for the functionalization of substrates that contain only one hydroxyl group, it would have even greater application if it could be used for molecules that have several hydroxyl groups. Whether selectivity can be obtained in such systems remains to be seen. Nonetheless, the authors’ innovative transformation sheds light on how common chemical groups can be used to direct C–H functionalization reactions, and provides a new and efficient way to prepare 1,3-diol units in complex organic molecules. ■

Danielle M. Schultz and John P. Wolfe are in the Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055, USA.
e-mail: jpwolfe@umich.edu

1. Simmons, E. M. & Hartwig, J. F. *Nature* **483**, 70–73 (2012).
2. Wencel-Delord, J., Dröge, T., Liu, F. & Glorius, F. *Chem. Soc. Rev.* **40**, 4740–4761 (2011).
3. Newhouse, T. & Baran, P. S. *Angew. Chem. Int. Edn* **50**, 3362–3374 (2011).

4. Chen, M. S. & White, M. C. *Science* **318**, 783–787 (2007).
5. Lyons, T. W. & Sanford, M. S. *Chem. Rev.* **110**, 1147–1169 (2010).
6. Engle, K. M., Mei, T.-S., Wasa, M. & Yu, J.-Q. *Acc. Chem. Res.* <http://dx.doi.org/10.1021/ar200185g> (2011).
7. Huang, C., Ghavtadze, N., Chattopadhyay, B. & Gevorgyan, V. *J. Am. Chem. Soc.* **133**, 17630–17633 (2011).

8. Simmons, E. M. & Hartwig, J. F. *J. Am. Chem. Soc.* **132**, 17092–17095 (2010).
9. Plé, K., Chwalek, M. & Voutquenne-Nazabadioko, L. *Eur. J. Org. Chem.* **2004**, 1588–1603 (2004).
10. García-Granados, A., López, P. E., Melguizo, E., Parra, A. & Simeó, Y. *J. Org. Chem.* **72**, 3500–3509 (2007).

MICROFLUIDICS

Analog-to-digital drug screening

Current methods for screening libraries of compounds for biological activity are rather cumbersome, slow and imprecise. A method that breaks up a continuous flow of a compound’s solution into droplets offers radical improvements.

ROBERT C. R. WOOTTON & ANDREW J. DEMELLO

Finding new leads for drug discovery increasingly depends on high-throughput methods that allow efficient biological screening of chemical libraries. Reporting in *Proceedings of the National Academy of Sciences*, Miller *et al.*¹ describe a method for screening large numbers of compounds — thousands or more — for their activity as enzyme inhibitors. This protocol generates immense amounts of high-precision dose-response data in ultrashort times of just a few minutes per compound. The authors’ approach could vastly improve the efficiency of current screening processes, which represent a major bottleneck for drug-discovery programmes.

Conventional routes to screening compounds typically begin with a qualitative primary screen of a diverse compound library to identify which compounds are active at a biological target. This is supplemented by a dose-response analysis of selected active compounds, to relate the target’s activity to each compound’s concentration. Unfortunately, dose-response screening of libraries is a time-consuming and complex procedure, with many potential pitfalls. Not least of these is the amount of time required to gather sufficient data to probe complex pharmacology, and the large number of false positives and false negatives generated from standard assays².

Library screening to find enzyme inhibitors in particular can be fraught with difficulties because of the high number of compounds that must be screened and the rather cumbersome methods used — typically involving robots running assays in tiny wells on plates. In practice, every compound in a primary screen is normally assayed at a single concentration; the vast majority of candidates are almost inevitably inactive and are thus eliminated. The first screen is then followed by a more complex,

dose-dependent assay of the remaining candidates. But although secondary assays are more detailed, they are frequently run at only a limited number of compound concentrations, because they face similar constraints to the primary screen³.

Assays based on microfluidics — the precise control of fluids constrained within sub-millimetre-scale channels — could transform high-throughput screening. Microfluidic technology is being increasingly used in chemistry and biology because of its ability to rapidly perform complex analytical procedures on minute sample volumes, with greater efficiency than traditional macroscale approaches⁴. Such systems normally operate either by passing a continuous flow of a solution through channels, or by breaking a flow into segments or droplets.

Continuous-flow systems offer some advantages over those involving segmented flow — for example, they allow chemical gradients to be established in a flow. But they are also hampered by sample dispersion, which spreads a compound sample out from an initial, focused volume to a wider, more dilute region. Continuous-flow systems also suffer from residence-time distributions, in which some parts of a sample move faster than others, thus adding to sample spreading and dilution. Segmented (or droplet-based) flows, on the other hand, offer excellent control over sample dispersion and residence time, although gradients are more difficult to establish.

Miller *et al.*¹ have exploited key features of both continuous- and segmented-flow microfluidics to develop a robust system capable of generating high-quality dose-response data in screens for enzyme inhibitors. At the heart of their approach is the generation of a concentration gradient over time in a flow of a solution of the compound using Taylor–Aris dispersion (Fig. 1a) — a phenomenon that occurs in continuous flows as a result of a

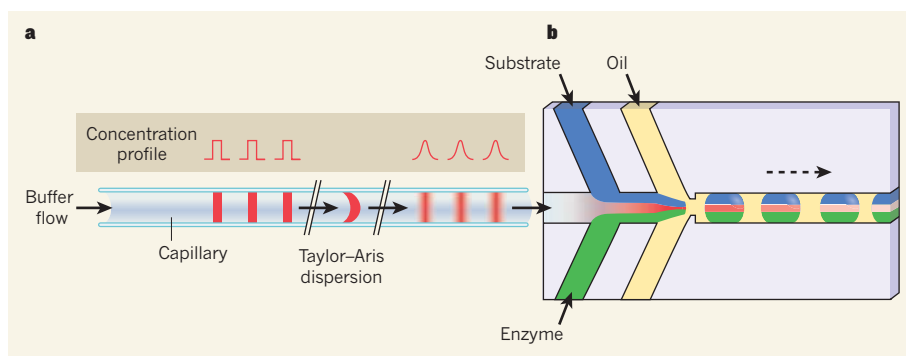


Figure 1 | High-resolution dose-response screening. Miller *et al.*¹ report a microfluidics system that enables the activity of enzyme inhibitors to be measured at thousands of different concentrations in just a few minutes. **a**, In their system, solutions of inhibitors (red) are injected into a continuous flow of buffer in a capillary tube. By means of a phenomenon known as Taylor-Aris dispersion, the concentration profile of each injected solution is transformed into a smooth pulse. **b**, Solutions of the targeted enzyme and of an enzyme substrate are introduced into the flow, which is then broken into droplets by the addition of an immiscible oil. Each droplet thus contains a slightly different concentration of the inhibitor. By measuring enzyme inhibition within each droplet, a dose-response curve can be plotted for the inhibitor. (Figure adapted from ref. 1.)

combination of molecular diffusion and the variation of flow velocity over the cross-section of a microchannel. The authors segment this flow into a series of droplets dispersed in a continuous flow of oil (Fig. 1b), in a process similar to the analog-to-digital conversion of an electrical signal.

Each droplet has a volume of just 140 picolitres (1 picolitre is 10^{-12} litres), and defines a unique reactor containing a concentration of the candidate inhibitor that differs from the concentrations in droplets preceding or following it. No dispersion of the dissolved compound can occur between droplets, and so the concentration gradient established in the original continuous flow is preserved. By incorporating the target enzyme and its substrate within each droplet and measuring the enzyme inhibition by the candidate, the authors were able to produce dose-response curves containing 10,000 data points for a compound from a single sample of solution, within a few minutes.

The large amount of data extracted from each experiment means that the resulting dose-response curves can be analysed with high precision using complex kinetic models. This greatly reduces the occurrence of false negatives and false positives, and generates much more helpful data than are produced by currently available secondary assays. Given that the authors' microfluidic device can screen each compound so effectively within two to three minutes, their approach also opens the way to more thorough and effective primary screening.

As Miller *et al.* highlight¹, their system still has some issues that must be resolved, such as ensuring that each droplet remains in the microfluidic system long enough for enzyme inhibition to occur. The authors currently achieve this by incorporating channels (known as delay lines) in which the flow is slowed. Although these lines allow reactions to be assayed over a period of 3.5 minutes,

they also introduce residence-time distributions because of droplet disordering. This is clearly problematic, especially when screening slower enzymatic processes, but could be overcome by changing delay-line geometry and channel lengths. Another concern is that the range of inhibitor concentrations that can be accessed by the authors' system is rather

limited. However, this issue could easily be resolved by incorporating passive droplet-dilution modules³ — devices that would dilute each sample-containing droplet into a stream of new, more-dilute droplets.

Despite these limitations, the impact of Miller and colleagues' approach on the drug-discovery process will be considerable. By enabling large amounts of dose-response information to be obtained from small sample volumes, it could facilitate the direct, quantitative dose-response screening of whole libraries in short times, eliminating the need for second-pass screening of active compounds. ■

Robert C. R. Wootton and Andrew J. deMello are at the Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology Zurich, CH-8093 Zurich, Switzerland. e-mails: andrew.demello@chem.ethz.ch; robert.wootton@chem.ethz.ch

1. Miller, O. J. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 378–383 (2012).
2. Brian, K. S. *Drug Discov. Today* **11**, 607–615 (2006).
3. Inglese, J. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 11473–11478 (2006).
4. Hong, J., Edel, J. B. & deMello, A. J. *Drug Discov. Today* **14**, 134–146 (2009).
5. Niu, X., Gielen, F., Edel, J. B. & deMello, A. J. *Nature Chem.* **3**, 437–442 (2011).

TRANSLATIONAL MEDICINE

Primed for resistance

A drug for treating melanoma is ineffective in colorectal cancers that have the same causative mutation. Studies of how cells adapt to the drug reveal why this is so, and suggest combination therapies that may be more effective. SEE LETTER P.100

DAVID B. SOLIT & PASI A. JÄNNE

Genetic mutations that disrupt fundamental cellular processes, such as cell division and differentiation, can lead to cancer. Mutations in the *BRAF* gene, which codes for the kinase protein BRAF, are one such example, and are detected in multiple cancers. In patients with advanced *BRAF*-mutant melanomas, treatment with the *BRAF*-inhibiting drug vemurafenib significantly prolongs survival¹. However, vemurafenib is ineffective² in patients with colorectal cancers that harbour the same *BRAF* mutation — *BRAF*(V600E). On page 100 of this issue, Prahallad *et al.*³ provide an explanation for this lack of efficacy by showing that, in colorectal cancer cells, drug treatment leads to rapid activation of a protein that counteracts the drug's *BRAF*-inhibitory activity.

Prahallad and colleagues identified this molecular basis for resistance by screening for genes for which inhibition of expression

leads to enhanced sensitivity of human colorectal cancer cells to vemurafenib. Their screen assessed all 518 human kinase-encoding genes and 17 additional kinase-related genes. Among the most potent 'hits' they identified was the gene coding for the epidermal growth factor receptor (*EGFR*). To test the biological significance of this finding, the authors treated *BRAF*(V600E)-mutant colorectal cancer cell lines with vemurafenib, and found that the drug induced hallmarks of *EGFR* activation.

EGFR acts earlier than *BRAF*, and other *RAF* proteins, in a cell signalling pathway called the *ERK* pathway. The authors hypothesize that inhibition of *BRAF* by vemurafenib relieves a negative feedback loop that keeps *EGFR* inactive. *EGFR* activates a protein called *RAS*, which, when activated, has been shown to cause vemurafenib resistance by inducing the formation of dimers of *RAF* proteins, against which the drug is ineffective (Fig. 1). *EGFR* also activates other cell signalling pathways,

such as the PI3K/AKT pathway, and EGFR-mediated activation of these pathways may also contribute to vemurafenib resistance.

The identification of EGFR activation as the basis for vemurafenib resistance in colorectal cancer cells is consistent with a model of resistance that has emerged from studies of patients with melanoma who acquire resistance to this drug^{4–6}. In *BRAF*-mutant melanomas, RAS activity is low, and consequently RAF dimer formation is low. Vemurafenib effectively inhibits the monomeric forms of BRAF⁶. However, perturbations that increase RAS activation (such as *RAS* mutation or activation of upstream kinases) promote the formation of RAF dimers, leading to drug resistance^{4–6}.

In contrast to melanoma cells, EGFR is ubiquitously expressed in *BRAF*-mutant colorectal cancers, although it is inactive before drug treatment. Exposure to vemurafenib results in rapid EGFR activation and immediate drug resistance. EGFR-expressing colorectal cancer cells are thus 'primed' for resistance to BRAF inhibitors. This model of vemurafenib resistance is supported by another recent study⁷, which reports that vemurafenib treatment leads to increased EGFR-mediated RAS activity in *BRAF*-mutant colorectal cancer cells. This can be blocked by concurrent treatment with an EGFR inhibitor.

Drugs that target EGFR, such as the monoclonal antibody cetuximab, are ineffective in patients with advanced *BRAF*-mutant colorectal cancer⁸. However, when Prahallad and colleagues³ treated mice bearing grafts of human *BRAF*-mutant colorectal cancers with vemurafenib and cetuximab, the combination was significantly more effective than either drug alone. These results imply that EGFR and BRAF inhibitors, despite being ineffective in patients with *BRAF*-mutant colorectal cancer when used alone, may be more successful when administered in combination.

Prahallad and colleagues' paper complements a series of recent publications that have enhanced our understanding of the seemingly idiosyncratic activity of BRAF inhibitors in patients. Specifically, vemurafenib inhibits ERK-pathway activity in *BRAF*-mutant tumour cells, but paradoxically activates ERK signalling in normal cells and in tumour cells expressing normal BRAF⁹. The drug thus inhibits ERK activity in a *BRAF*-mutant-specific manner, an atypical property that accounts for its unusual toxicity profile. Approximately one-third of patients treated with vemurafenib develop a skin rash that differs markedly from the rashes associated with drugs that inhibit ERK signalling. Some vemurafenib-treated patients also develop hyperproliferative skin lesions — including keratoacanthomas and cutaneous squamous-cell carcinomas — that do not harbour *BRAF* mutations¹. The emergence of these skin lesions is explained by two recent studies^{10,11} that found that most of the lesions harbour

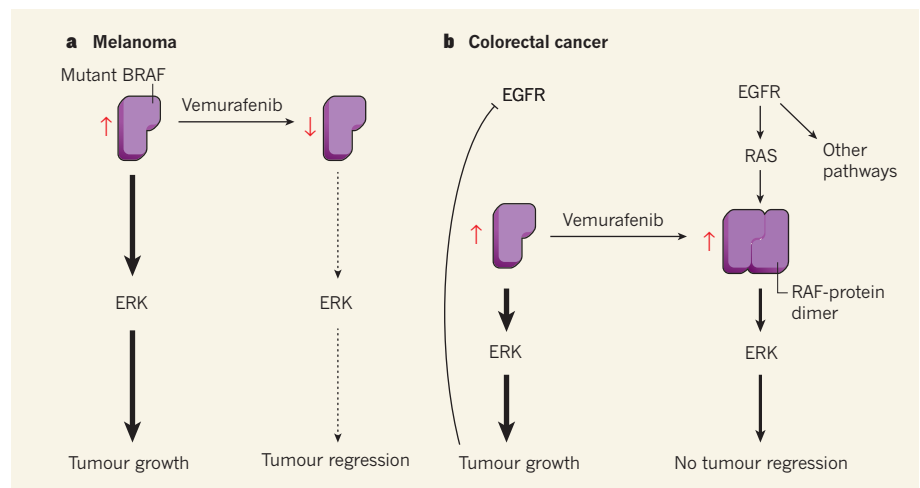


Figure 1 | Intrinsic drug resistance in colorectal cancer cells. **a, b,** Cancers with mutations in the *BRAF* gene show hyperactivity (represented by up arrows) of BRAF, a protein in the ERK cellular signalling pathway. This induces overactivation of the pathway (shown by bold arrows), which promotes tumour growth. **a,** *BRAF*-mutant melanomas are sensitive to vemurafenib, a drug that inhibits mutant BRAF. In these sensitive cells, vemurafenib inhibits ERK-pathway activity, leading to tumour regression. **b,** Colorectal cancer cells with the same mutation are resistant to vemurafenib. Prahallad *et al.*³ show that this drug resistance arises as a result of rapid activation of the epidermal growth factor receptor (EGFR), which acts earlier than BRAF in the ERK pathway. EGFR is broadly expressed in colorectal cancer cells but is kept in an inactive state in *BRAF*-mutant cells by an ERK-dependent negative feedback loop. Vemurafenib inhibition of BRAF relieves this negative feedback, resulting in rapid EGFR activation. EGFR then activates the protein RAS, which promotes dimerization of BRAF or other RAF proteins. As RAF dimers are resistant to vemurafenib, activation of EGFR abrogates the inhibitory effects of the drug.

mutations in the genes encoding RAS proteins. Animal experiments have indicated that, as a result of the elevated RAS activity in these skin lesions, vemurafenib activates ERK signalling, which leads to cellular hyperproliferation.

Taken together, these findings^{3–11} have broad clinical implications. They clearly justify studies combining inhibitors of EGFR and BRAF inhibitors in patients with *BRAF*-mutant colorectal cancer. Furthermore, as vemurafenib activates ERK signalling and EGFR inhibitors have the opposite effect, this drug combination may prove less toxic than either drug alone, a result similar to the reduced toxicity observed in studies combining inhibitors of MEK (another ERK-pathway protein) and BRAF¹².

Prahallad and colleagues' findings also highlight the potential for differences in efficacy of kinase-targeted therapies in tumours that share the same mutation but arise in different tissue types. For example, because EGFR is broadly expressed in epithelial cells, carcinomas may, as a group, be intrinsically less sensitive to BRAF inhibition than melanomas, which occur in cells of a different (neural crest) origin from carcinomas. This hypothesis awaits clinical validation in ongoing trials of BRAF inhibitors.

As efforts to characterize individual tumours on a molecular basis accelerate in the clinic, with the goal of personalizing cancer therapy, there will be a strong desire to treat patients who have a *BRAF*-mutant tumour with a BRAF inhibitor, regardless of the site of tumour origin. The findings reported by Prahallad *et al.* suggest that this approach may

sometimes prove ineffective. Ideally, such efforts at personalized medicine should proceed within the context of clinical trials and should include molecular analysis of both pre- and post-treatment tumour samples, because such studies may aid the identification of additional cell-type-specific factors that affect drug responses and thereby facilitate the design of combination therapies. ■

David B. Solit is in the Department of Medicine and the Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. **Pasi A. Jänne** is at the Lowe Center for Thoracic Oncology, Dana Farber Cancer Institute, Boston, Massachusetts 02215, USA. e-mails: solitd@mskcc.org; pjanne@partners.org

1. Chapman, P. B. *et al.* *N. Engl. J. Med.* **364**, 2507–2516 (2011).
2. Kopetz, S. *et al.* *J. Clin. Oncol.* **28** (ASCO Meet. Abstr. Pt 1), abstr. 3534 (2010).
3. Prahallad, A. *et al.* *Nature* **483**, 100–103 (2012).
4. Poulikakos, P. I., Zhang, C., Bollag, G., Shokat, K. M. & Rosen, N. *Nature* **464**, 427–430 (2010).
5. Nazarian, R. *et al.* *Nature* **468**, 973–977 (2010).
6. Poulikakos, P. I. *et al.* *Nature* **480**, 387–390 (2011).
7. Corcoran, R. B. *et al.* *Cancer Discov.* <http://dx.doi.org/10.1158/2159-8290.CD-11-0341> (2012).
8. Di Nicolantonio, F. *et al.* *J. Clin. Oncol.* **26**, 5705–5712 (2008).
9. Joseph, E. W. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 14903–14908 (2010).
10. Oberholzer, P. A. *et al.* *J. Clin. Oncol.* **30**, 316–321 (2012).
11. Su, F. *et al.* *N. Engl. J. Med.* **366**, 207–215 (2012).
12. Infante, J. R. *et al.* *J. Clin. Oncol.* **29** (ASCO Meet. Abstr. Pt 1), abstr. CRA8503 (2011).

KNOWLEDGE

You must remember this.

BY JOHN FRIZELL

No human could have grasped the squid's name. Human eyes could not distinguish the differences in shades of colour or register the intervals at which they changed to define the unique pattern that was his name. The squid was concentrating hard because he was holding two conversations at once, one deliberately misleading, the other closer to the truth, as he glided through the deep ocean, his mantle pulsing gently, powering him with puffs of water.

On his right side the flowing patterns of colour were attempting to persuade a female to mate with him. He was being as persuasive as possible without exaggerating; females were very good at picking up on male lies, and like all female giant squid she was considerably bigger than him, or any adult male.

On his left he was chatting to another male about recent sightings of Loud Shadows, the huge predators that were the only things in the ocean that attacked giant squid.

"Our hatched memories are not right," he said.

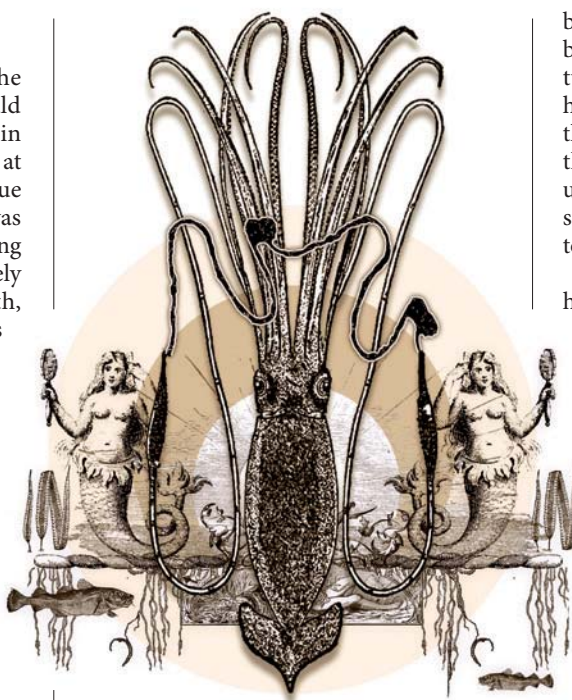
A squid was lucky to live for more than 80 or 90 lunar cycles. But everyone had memories with which they were hatched and that went back hundreds of millions of cycles.

"It was much worse back then," said the other male. "The ocean was infested with Shadows. But the light-songs we get from others say that thousands of cycles ago, they began to disappear. The legend is that they were hunted by a surface predator. But it is so hard to know. We have to rely on stories passed from one generation to the next, on memories of memories. But it seems that the surface predator has ended its hunt and the Shadows are returning."

Then, like a horror called up by saying its name, a Shadow's ranging pulse swept over them. All three went dark.

The squid's hearts pounded faster, and he shot off on a jet of water, veering away from the source of the sound. There was another pulse and then the Killing Sound, the noise that the Shadows used to disorient their victims while their terrible long jaw with its sharp conical teeth dismembered its prey. But the sound was not directed at him.

The female lit up in a display of rage and defiance. She was right in front of the Shadow and it hung there in the water,



pouring sound energy into her. She lashed out with her arms and wrapped them around its massive head, locking its jaw shut. If she could hang on long enough, she might live. The Shadows were not proper sea creatures: they had to return to the surface at regular intervals and could not remain at depth for long. Even so, her chances were not good. She was almost as long as a Shadow, and the sharp cutting rings of her suckers would be ripping into its skin, but that rarely made them give up. The monster was a solid mass of muscle, many times thicker than the thickest tentacle.

The squid swam away at full speed. Shadows sometimes hunted in groups. When he slowed, he was far away and as safe as he could be in the ocean, where an attack could come at any time. He was ravenous from the energy expenditure and did a bit of hunting of his own, caught a fish and pushed it into his beak where his radula quickly shredded it and fed it into his stomach. He could feel the energy spreading through his body.

His rival had got one thing very right. Because they had to rely on memories of memories, the past was hard to know. And

➔ **NATURE.COM**
Follow Futures on
Facebook at:
go.nature.com/mtoodm

if you did not know the past, how could you learn from it? How could knowledge be preserved other than

by memory? The preservation of memory became an obsession, doomed in the featureless ocean wilderness. In desperation he had tried to dive to the bottom where there might be something he could use, but the deep was a place of strange smells and unsettling cold, a hostile environment that sucked energy out of him and made it hard to move.

Cycles went by. All too soon he found himself more than 80 cycles old. Squid do not become feeble with age or decline slowly — they retain their vigour to the end — but when the end comes, it comes quickly. He could feel his digestive system shrivelling. He would be lucky to last another cycle.

He was hunting for a snack when he made his breakthrough.

As he glided down on his prey he realized it was not a fish: it was Transparent Death, one of the unbreakable filaments that had started appearing hundreds of cycles ago. He banked away automatically, then circled back, thinking furiously. To preserve knowl-

edge independently of memory, it needed to be stored outside the body. Here was something that would last for thousands of cycles, perhaps millions. He flexed his tentacles. Tentacles were for catching and eating, occasionally for fighting if subterfuge and diplomacy failed, but if they could hold a fish, they could hold other things as well.

He approached the drifting strand as carefully as if it were a sleeping Shadow, caught an end and swam slowly away, trailing it so it could not wrap around him. He grasped it farther down with an arm and then caught the free end with his two tentacles, formed it into a loop and pulled on it. It started to tighten around a tentacle and he whipped it out before it was caught. He tried again. Finally he had the strand crossing over itself and grasping itself to form a nodule. He moved a bit down the filament and tried again. The second knot was easier. If he allowed each nodule to represent 'yes' and the absence of a nodule to represent 'no', he could record information.

He set off, carefully trailing the line, to show his discovery to others before it was too late. ■

John Frizell was trained in biochemistry and works in ocean conservation for Greenpeace. In his spare time he walks, builds robots and writes short stories.

JACEY

Structural and functional conservation of key domains in InsP₃ and ryanodine receptors

Min-Duk Seo^{1†*}, Saroj Velamakanni^{2*}, Noboru Ishiyama¹, Peter B. Stathopoulos¹, Ana M. Rossi², Samir A. Khan², Philippa Dale², Congmin Li³, James B. Ames³, Mitsuhiro Ikura¹ & Colin W. Taylor²

Inositol-1,4,5-trisphosphate receptors (InsP₃Rs) and ryanodine receptors (RyRs) are tetrameric intracellular Ca²⁺ channels¹. In each of these receptor families, the pore, which is formed by carboxy-terminal transmembrane domains, is regulated by signals that are detected by large cytosolic structures. InsP₃R gating is initiated by InsP₃ binding to the InsP₃-binding core (IBC, residues 224–604 of InsP₃R1)² and it requires the suppressor domain (SD, residues 1–223 of InsP₃R1)^{2–8}. Here we present structures of the amino-terminal region (NT, residues 1–604) of rat InsP₃R1 with (3.6 Å) and without (3.0 Å) InsP₃ bound. The arrangement of the three NT domains, SD, IBC-β and IBC-α, identifies two discrete interfaces (α and β) between the IBC and SD. Similar interfaces occur between equivalent domains (A, B and C) in RyR1 (ref. 9). The orientations of the three domains when docked into a tetrameric structure of InsP₃R¹⁰ and of the ABC domains docked into RyR⁹ are remarkably similar. The importance of the α-interface for activation of InsP₃R and RyR is confirmed by mutagenesis and, for RyR, by disease-causing mutations^{9,11,12}. Binding of InsP₃ causes partial closure of the clam-like IBC, disrupting the β-interface and pulling the SD towards the IBC. This reorients an exposed SD loop ('hotspot' (HS) loop) that is essential for InsP₃R activation⁷. The loop is conserved in RyR and includes mutations that are associated with malignant hyperthermia and central core disease^{9,11,12}. The HS loop interacts with an adjacent NT, suggesting that activation re-arranges inter-subunit interactions. The A domain of RyR functionally replaced the SD in full-length InsP₃R, and an InsP₃R in which its C-terminal transmembrane region was replaced by that from RyR1 was gated by InsP₃ and blocked by ryanodine. Activation mechanisms are conserved between InsP₃R and RyR. Allosteric modulation of two similar domain interfaces within an N-terminal subunit reorients the first domain (SD or A domain), allowing it, through interactions of the second domain of an adjacent subunit (IBC-β or B domain), to gate the pore.

The essential role of the SD in linking InsP₃ binding to InsP₃R gating highlights the need to define the structural consequences of InsP₃ binding to the N-terminal region (NT, residues 1–604 of InsP₃R1) (Supplementary Fig. 1). Because our attempts to crystallize the NT yielded poorly diffracting crystals, we expressed a Cys-less form of the NT (NT(Cys-less)). Native and Cys-less forms of the NT and IBC behaved indistinguishably (Supplementary Fig. 2 and Supplementary Tables 1 and 2), but NT(Cys-less) provided crystals with diffraction that was much improved (Supplementary Table 3). We determined crystal structures of NT(Cys-less) with (3.6 Å) and without (3.0 Å) InsP₃ bound, showing three domains: the SD, IBC-β (residues 224–436) and IBC-α (residues 437–604) (Fig. 1a). The structures of these domains were nearly identical to those of isolated native SD and IBC^{2,3} (Supplementary Fig. 3).

The SD, IBC-β and IBC-α form a triangular structure, with the SD behind the InsP₃-binding site (Fig. 1a). The SD interacts with the IBC

through two interfaces, one with IBC-β (β-interface) and another with IBC-α (α-interface). A 3₁₀-like turn between the last strand of the SD and the first strand of IBC-β positions the IBC relative to the SD (Supplementary Fig. 4e). Within this connecting turn, a salt bridge (between Lys225 and Asp228) stabilizes the backbone conformation and thus positions the residues that form the β-interface. These interactions in the connecting turn and β-interface are augmented by a network of hydrophobic interactions within IBC-β (Fig. 1b). The α-interface forms a long 'Velcro'-like structure that also involves a network of hydrophobic and electrostatic interactions (Fig. 1c). Intimate hydrophobic interactions between Val33, and to a lesser extent Leu32, from the SD; and Val452, Phe445, Ala449 and Leu476 from IBC-α are supported by bidentate salt bridges between Arg54 and Lys127 in the SD and Asp444 in IBC-α (Fig. 1c). The Val33Lys mutation at the α-interface almost abolished inhibition of InsP₃ binding by the SD^{3,4} and reduced the probability of the channel being open⁴. This confirms the functional importance of the α-interface. Mutation of neighbouring residues that contribute less to the α-interface (Leu32Lys, Asp34Lys, Arg36Glu, Lys127Glu) had less effect on InsP₃ binding, and mutation of residues that do not contribute to the interface (Asp35Lys, Lys52Glu) had no effect (Supplementary Table 4)^{3,4}. Hydrophobic and electrostatic interaction networks at the α- and β-interfaces contribute to a buried surface between the SD and IBC (~2,040 Å²) that forms a hub connecting InsP₃ binding to channel activation.

The structure of the NT is very similar to that of the N-terminal region of RyR1 (ref. 9). The three NT domains of InsP₃R1 (SD, IBC-β and IBC-α) can be individually superposed to corresponding domains of RyR1 (A, B and C) (Supplementary Fig. 3). Furthermore, the relative orientation of the three domains within each N-terminal structure is nearly identical (Fig. 1d). Mutation of Tyr167Ala, located on an exposed loop of the SD opposite the IBC interfaces (HS loop¹¹, residues 165–180, boxed in Fig. 1d), attenuates InsP₃-evoked Ca²⁺ release⁸, and Ca²⁺, a co-regulator of InsP₃R¹³, causes the loop to become accessible¹⁴. The disease-associated HS loop of RyR1 (ref. 11) sits at the same location within the ABC structure⁹ (Fig. 1d), and a mimetic peptide causes RyR2 to become leaky¹⁵. Furthermore, the backbone and side-chain conformation of this loop region superposes well in the two receptors (Fig. 1e). The HS loop provides a critical link between InsP₃ binding and gating.

The domain interfaces of the NT of InsP₃R1 and the ABC domains of RyR1 are also similar. The bidentate salt bridges at the InsP₃R1 α-interface between Arg54 and Lys127 in the SD and Asp444 in IBC-α are preserved in RyR1 ABC, albeit in a reversed-charge manner between Asp40 and Asp61 in the A domain and Arg402 in the C domain (Supplementary Fig. 4a). In RyR1, mutation of these residues (Arg402Cys, Asp61Asn) is associated with malignant hyperthermia and central core disease⁹, suggesting that disruption of the interaction perturbs RyR gating, as it does for InsP₃R. The structural similarities

¹Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Ontario M5G 1L7, Canada. ²Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1PD, UK. ³Department of Chemistry, University of California, Davis, California 95616, USA. [†]Present address: College of Pharmacy, Ajou University, Suwon 443-749, Korea.

*These authors contributed equally to this work.

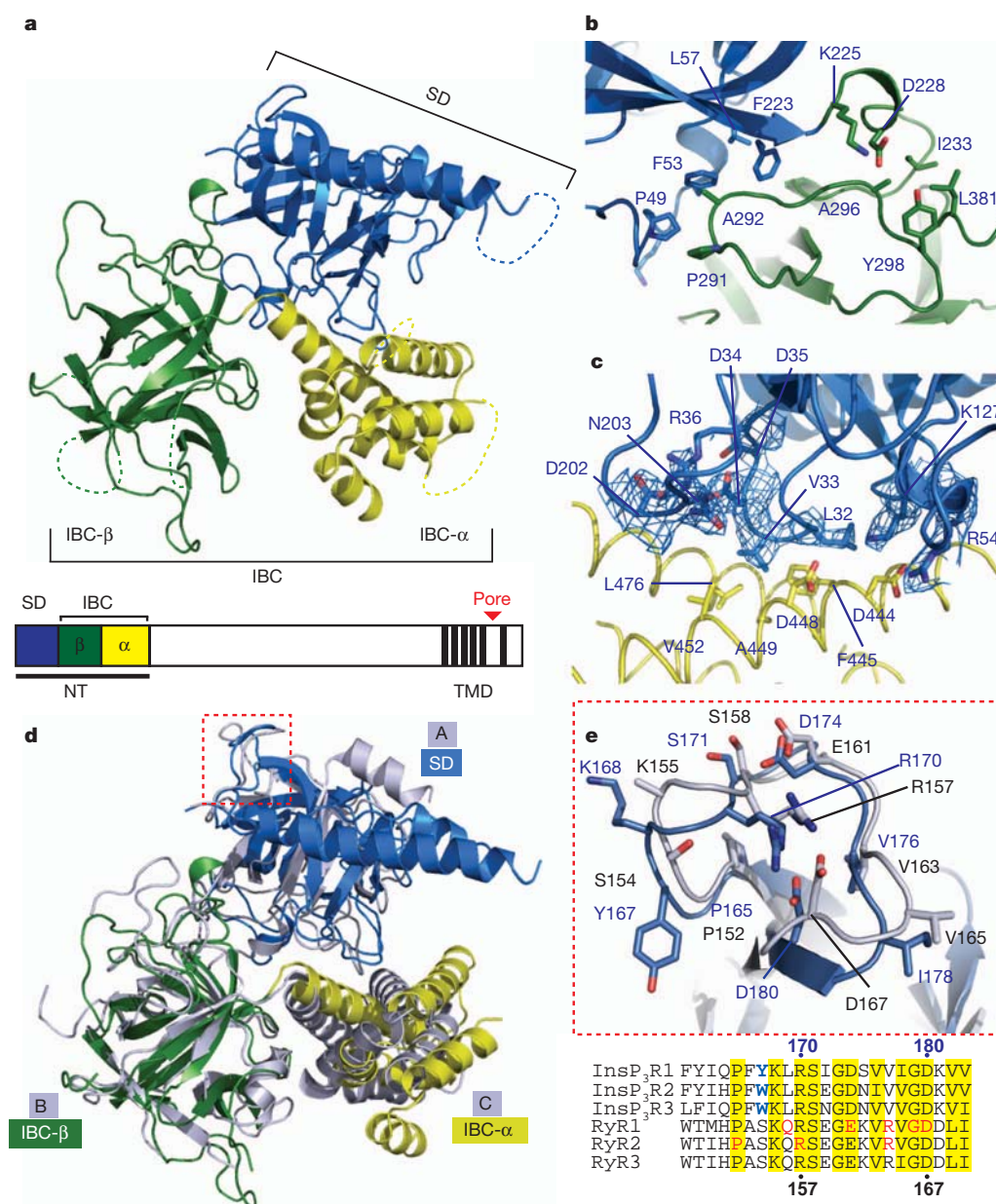


Figure 1 | Structure of the NT region of InsP₃R1 without InsP₃ bound.

a, Structure of NT(Cys-less) at 3 Å resolution showing SD (blue), IBC-β (green) and IBC-α (yellow). Dashed lines show invisible regions in the electron density. Positions of the three domains within a single InsP₃R subunit are shown. TMD, transmembrane domain. **b**, **c**, Interfaces between SD-IBC-β (β-interface) and SD-IBC-α domains (α-interface), with the hydrophobic core boxed and the 2F_o - F_c electron density map of key residues (contoured at 1.0σ) shown as mesh) (**c**). **d**, Superposition of apo-NT(Cys-less) and the ABC domain of RyR1

(grey)⁹ structures by overlaying IBC-β and the RyR1 B domain. The HS loop in RyR1 and corresponding region in InsP₃R1 are highlighted (red box). A, B and C indicate domains of RyR1. **e**, Close-up views of HS regions of InsP₃R1 (blue) and RyR1 (grey) with conserved residues depicted as sticks. Structure-based DaliLite alignment of rat InsP₃Rs and rabbit RyRs shows conserved residues (yellow), RyR1 disease-associated mutations (red), and hydrophobic residues that are implicated in activation of InsP₃R (blue)⁸.

also extend to the β-interface of InsP₃R1 and the corresponding A-B interface in RyR1 (Supplementary Fig. 4b, d). Interestingly, RyR1 forms an additional salt bridge at the A-C interface between Arg45 and Asp447, which is not conserved in InsP₃R1 (Supplementary Fig. 4c).

Our structures of NT(Cys-less) with and without InsP₃ bound, together with the structure of the InsP₃-bound IBC² (Supplementary Fig. 5), reveal the structural changes that are evoked by InsP₃ (Fig. 2). Side chains of nine residues become organized around InsP₃ (Supplementary Fig. 5a), and the domain orientation angle between IBC-β and IBC-α is reduced (by ~8°) after InsP₃ binding (Fig. 2 and Supplementary Fig. 5a). This InsP₃-evoked 'clam closure', which is consistent with earlier predictions¹⁶ and small-angle X-ray scattering¹⁷, causes the distance across the entrance to the InsP₃-binding pocket to decrease

(Supplementary Fig. 5b, c). A similar agonist-evoked domain closure occurs in some glutamate receptor channels¹⁸. The SD and IBC remain associated after closure of the IBC (Fig. 2). InsP₃ binding hardly changes the interactions across the extensive α-interface, but at the β-interface the SD residues move away from IBC-β (Supplementary Fig. 5d-f). With the SD glued to IBC-α by the α-interface, and the β-interface serving as a lubricant, InsP₃ binding causes the SD to twist (by ~9°) and move closer to the top of the IBC (Fig. 2). This causes an amplified translational movement of the conserved HS loop in the SD (Supplementary Fig. 5g). While our work was under review, 3.8 Å structures of apo- and InsP₃-bound NT that were derived from a single crystal grown in excess InsP₃ were published, showing similar InsP₃-induced allostery in the interfaces between domains¹⁹. This confirms

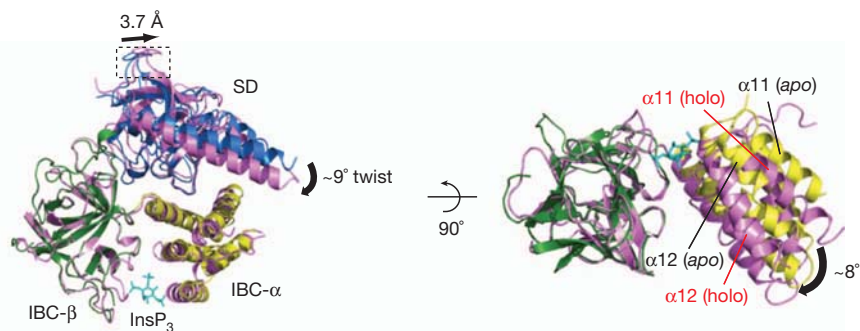


Figure 2 | InsP₃-evoked conformational changes. Superposition of *apo*-NT(Cys-less) (SD, blue; IBC-α, yellow; IBC-β, green) and InsP₃-bound NT(Cys-less) (3.6 Å resolution, magenta) by overlaying IBC-β domain. InsP₃ binding causes the SD to rotate towards the IBC accompanied by a swing that is approximately perpendicular to the IBC 'clam closure'. This twist is measured as the angular difference between the SD arm helices in the *apo*- and InsP₃-

bound states ($\sim 9^\circ$). Movement of the HS loop (shown in a dashed box) shows the distance between α -carbons of Tyr 167 (~ 3.7 Å). A view rotated 90° about the x axis is shown (right) with only IBCs represented. The interdomain (IBC-β and IBC-α) angular difference between the free and bound states is $\sim 8^\circ$. For further details of InsP₃ binding and its effects on the IBC and α - and β -interfaces see Supplementary Fig. 5.

our observations, but our higher resolution structures reveal more detail of the α - and β -interfaces that are associated with this conformational change (Supplementary Discussion).

Docking the ABC structure into cryo-electron microscopy maps of RyR1 showed that the N-terminal domains form a central ring at the top of the mushroom-like RyR1 (ref. 9). Rigid-body docking of our *apo*-NT(Cys-less) structure into a cryo-electron microscope 10 Å structure of a closed InsP₃R1 (ref. 10) reveals an arrangement that is remarkably similar to that of RyR1 with a high docking contrast (Fig. 3 and Supplementary Fig. 6). The three domains of the four NTs, which form the upper cytoplasmic surface of the mushroom-like InsP₃R, are arranged as four 'hillocks' around a central 'bowl'. This arrangement allows InsP₃ unrestricted access to the IBC from the side of the cap (Fig. 3), and it is consistent with accessibility studies and binding sites for regulatory proteins (Supplementary Fig. 6c and Supplementary Table 5). Within the tetrameric InsP₃R, the only contacts between NT subunits are through the critical HS loop of the SD and a flexible loop between β -strand 20– β -strand 21 (Supplementary Fig. 7) in IBC-β (Fig. 3c, d). The flexible loop is longer in RyR and it lies ~ 10 Å further from the neighbouring HS loop⁹ (Fig. 3c, d). In InsP₃R, the arm domains (residues 67–109) of each SD are the only NT structures that extend beyond the cap towards the pore (Fig. 3a), but these domains are neither essential for InsP₃R activation⁸ nor conserved in RyR^{9,11}.

The structural similarities between the N termini of RyR and InsP₃R prompted us to examine whether the domains are functionally interchangeable. In a chimaeric NT fragment comprising the A domain of

RyR2 and the IBC from InsP₃R1 (RyR2A-IBC), the A domain mimicked the SD by inhibiting InsP₃ binding (Fig. 4a, b). Mutations within the A-domain loop that forms the A–B interface in RyR⁹ or the equivalent InsP₃R loop in the SD attenuated this inhibition of binding (Fig. 4c, Supplementary Table 6 and Supplementary Fig. 8). InsP₃ stimulated Ca²⁺ release through InsP₃R1 or a chimaeric InsP₃R1 in which the SD was replaced by the A domain of RyR1 (RyR1A-InsP₃R1 (Fig. 4a, d and Supplementary Fig. 9). These InsP₃Rs were similarly expressed and they released similar fractions of the Ca²⁺ stores and had similar sensitivity to InsP₃ (Supplementary Table 7). Opening of native InsP₃R or RyR is restrained by interactions between cytosolic domains^{20,21}. It is therefore noteworthy that expression of InsP₃R1 or RyR1A-InsP₃R1 affected neither the Ca²⁺ content of the endoplasmic reticulum nor the Ca²⁺ leak from it (Supplementary Fig. 10), confirming that InsP₃R and RyR1A-InsP₃R1 have no detectable spontaneous activity. This demonstrates that the SD of InsP₃R can be functionally substituted by the A domain of RyR.

An InsP₃R1 in which residues downstream of transmembrane domain 1 were replaced by the equivalent region of RyR1 (InsP₃R1-RyR1) also responded to InsP₃ (Fig. 4a, e). Expression of InsP₃R1-RyR1 increased Ca²⁺ leakage from the endoplasmic reticulum, and this was reversed by ryanodine, which blocks the RyR pore²². However, the increased leak was insufficient to affect the steady-state Ca²⁺ content (Supplementary Fig. 10), suggesting that InsP₃R1-RyR1 has minimal spontaneous activity. Expression of InsP₃R1-RyR1 matched that of other InsP₃Rs, but cells expressing InsP₃R1-RyR1 were approximately 20-fold less sensitive to InsP₃ (Supplementary Table 7). Because

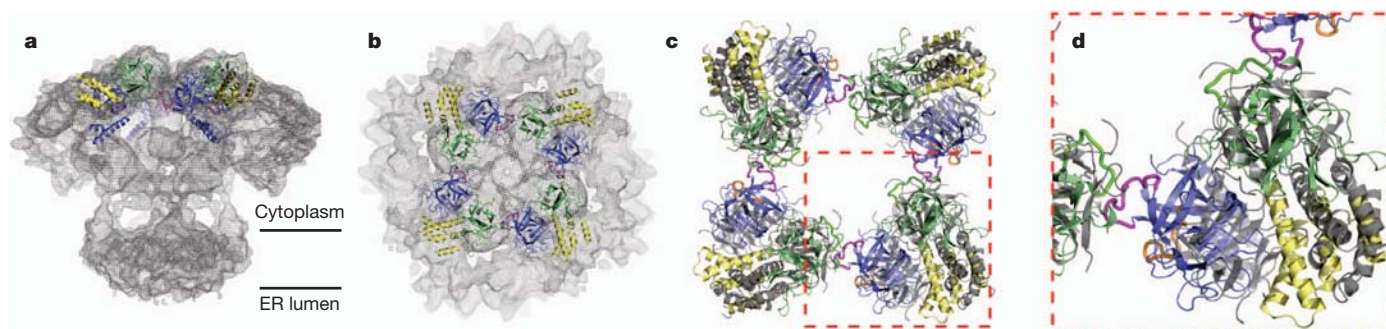


Figure 3 | Docking of the *apo*-NT(Cys-less) structure into the cryo-electron microscopy map of InsP₃R1. a, b, Side (a) and top (b) views of the *apo*-NT(Cys-less) structure docked into the cryo-electron microscopy map (grey mesh) of InsP₃R1 in a closed state¹⁰. Contour level corresponds to mass of InsP₃R1 tetramer of 1.3 megadaltons (MDa) (protein density $0.8 \text{ Da } \text{\AA}^{-3}$). Four molecules of the NT (SD, blue; IBC-β, green; IBC-α, yellow) are located at the top of the cytoplasmic portion of the InsP₃R1 tetramer. c, Dockings of the *apo*-

NT(Cys-less) (coloured as in a) and ABC⁹ (grey) structures into cryo-electron microscopy structures of InsP₃R1 (ref. 10) and RyR1 (ref. 9), respectively, are overlaid and presented to show only NT structures. HS loops of InsP₃R (magenta) and RyR (orange) are highlighted. d, Enlargement of boxed area in c. Locations of other binding sites within the NT of InsP₃R1 are shown in Supplementary Fig. 6.

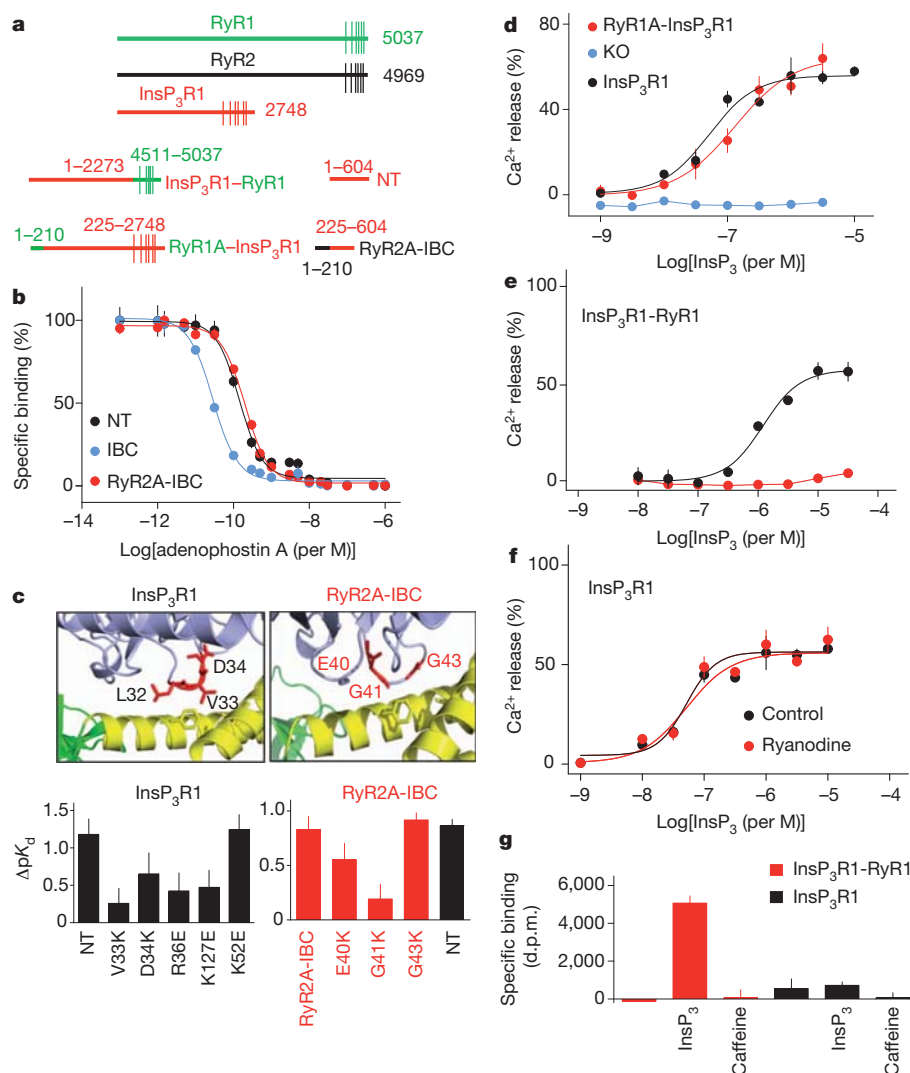


Figure 4 | Functional chimaeras of InsP₃R and RyR. **a**, Proteins used. **b**, Specific binding of ³H-InsP₃ in the presence of adenophostin A. **c**, Inhibition of ³H-InsP₃ binding to IBC by SD or A domain, and effects of mutations within equivalent loops. Affinities shown relative to IBC (ΔpK_d). The apo-NT(Cys-less) (left) and RyR2A-IBC homology (right) structures show key residues (red sticks) within the SD or A domain at the α-interface. The SD/A, IBC-β and IBC-α domains are shown in blue, green and yellow, respectively. **d**, Ca²⁺ release

from DT40 cells expressing InsP₃R1, RyR1A-InsP₃R1 or lacking InsP₃R (knockout (KO)). **e**, **f**, Effect of ryanodine (10 μM) on Ca²⁺ release from DT40 cells expressing InsP₃R1-RyR1 (**e**) or InsP₃R1 (**f**). Results (**d**–**f**) are percentages of ATP-dependent Ca²⁺ uptake. **g**, Specific ³H-ryanodine binding to membranes of DT40 cells expressing InsP₃R1 or InsP₃R1-RyR1 with caffeine (10 mM) or InsP₃ (1 μM). Nonspecific binding was 2245 ± 211 disintegrations per min (d.p.m.). Results (**b**–**g**) are means ± s.e.m., *n* ≥ 3.

the transmembrane domains minimally affect InsP₃ binding²³, this diminished response probably reflects a decrease in InsP₃ efficacy. The increased Ca²⁺ leakage and reduced efficacy of InsP₃ suggest that within InsP₃R1-RyR1, communication between the SD and channel are slightly less effective than in native InsP₃R. Nevertheless, it is remarkable that cytosolic domains of an InsP₃R should so effectively regulate the pore of a RyR when the two receptors share only modest sequence identity and differ in the number of residues separating the NT from transmembrane domains (Fig. 4a), and in the lengths and sequences of their C-terminal tails and the loops linking transmembrane domains (Supplementary Fig. 11).

Ryanodine (10 μM) had no effect on InsP₃R1 or RyR1A-InsP₃R1, but it abolished InsP₃-evoked Ca²⁺ release through InsP₃R1-RyR1 (Fig. 4e, f). Because ryanodine binds selectively to active RyR²⁴, ³H-ryanodine binding is stimulated by agonists of RyR, such as caffeine. Caffeine had no effect on specific ³H-ryanodine binding to InsP₃R1-RyR1, whereas InsP₃ stimulated it (Fig. 4g). InsP₃ therefore causes conformational changes to the channel of InsP₃R1-RyR1 that mimic those of native RyR in allowing binding of ³H-ryanodine. Ryanodine (≤10 μM) did not stimulate Ca²⁺ release through InsP₃R1-RyR1 (not

shown), suggesting that transmembrane domains alone may not mediate stimulation of RyR²⁵.

Conservation of structure–function relationships between InsP₃R and RyR (Figs 1–4) allows comparisons between them to suggest possible mechanisms of InsP₃R activation. For both receptors, gating requires conformational changes in the large cytoplasmic structures to pass to the transmembrane domains^{10,22}, but the N-terminal domains of InsP₃R and RyR are at least 60 Å from these transmembrane domains^{1,9,22} (Fig. 3a and Supplementary Fig. 6). Despite some evidence implicating direct interactions between the SD and the loop linking transmembrane domains 4 and 5 in gating InsP₃R (Supplementary Discussion), we suggest, and consistent with results from RyR^{20,26,27}, that additional cytosolic domains couple the NT to opening of the pore. The exposed HS loop in the SD (Figs 1d and 3c, d) (HS loop of RyR)^{9,11} is arranged similarly within the isolated N-terminal structures of InsP₃R and RyR (Fig. 1d) and it reorients after InsP₃ binding (Fig. 2). When the NT is docked into the InsP₃R structure¹⁰, the HS loop forms (with an exposed loop of IBC-β) the only interface between adjacent NT regions. However, the equivalent loop is displaced when the ABC is docked into the tetrameric RyR structure⁹ (Fig. 3c, d). InsP₃ binding closes the

clam-like IBC, disrupting the β -interface and reorienting the HS loop (Fig. 2 and Supplementary Fig. 5). We suggest that this disrupts interaction of the HS loop with a neighbouring NT to cause a coordinated rearrangement of the apical InsP₃R structure (Fig. 3). The open state of RyR1 is associated with outward movement of protein density in regions that match the locations of docked ABC structures^{9,22} and with larger movements of peripheral 'clamp domains' (refs 9, 22) that are absent from InsP₃R¹⁰. Movement of these apical domains in RyR is accompanied by rearrangements within regions that taper towards the pore²² and which, in InsP₃R, include the most flexible parts of its structure¹⁰. We suggest that similar rearrangements of the apical surface of InsP₃R and RyR couple to additional cytosolic domains by shared mechanisms to gate the pore of each channel.

METHODS SUMMARY

The NT residues 1–604 of rat InsP₃R1 in which all Cys residues were replaced by Ala (NT(Cys-less)) was expressed in *Escherichia coli* and purified. Crystals of NT(Cys-less) were grown by the hanging-drop vapour diffusion method in 0.1 M HEPES pH 7.0, 0.8–1.0 M (NH₄)₂SO₄ and 3% (v/v) trimethylamine N-oxide for *apo*-state crystals, or 0.1 M Na citrate (pH 6.0), 8% (w/v) polyethylene glycol 6000, 70 mM Li₂SO₄ and 3% dimethyl sulfoxide for InsP₃-bound crystals. Diffraction data were collected at 100 K on the 19-ID (*apo*-crystals) or 19-BM (InsP₃-bound crystals) beam lines at the Advanced Photon Source Synchrotron facility and processed with HKL2000 (ref. 28). Structures of *apo*-NT(Cys-less) at 3.0 Å resolution and InsP₃-bound NT(Cys-less) at 3.6 Å resolution were determined by molecular replacement using structures of the SD (Protein Data Bank code 1XZZ)³ and the IBC (1N4K)² as search models with the program Phaser²⁹. Iterative refinement and model building were performed with Refmac5 and Coot, respectively. Numbering of secondary structure motifs is in accord with Supplementary Fig. 7. Binding of ³H-InsP₃ or ³H-ryanodine to full-length InsP₃R1, chimaeras of InsP₃R1 and RyR, and to related NT fragments was defined using equilibrium-competition binding assays⁴. Functional properties of InsP₃R1 and chimaeras were characterized after stable expression in DT40 cells lacking endogenous InsP₃R⁴. A luminal Ca²⁺ indicator was used to record InsP₃-evoked Ca²⁺ release from the intracellular stores of permeabilized DT40 cells⁴.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 July; accepted 2 December 2011.

Published online 29 January 2012.

- Serysheva, I. (ed.) *Structure and Function of Calcium Release Channels*. (Academic Press, 2010).
- Bosanac, I. *et al.* Structure of the inositol 1,4,5-trisphosphate receptor binding core in complex with its ligand. *Nature* **420**, 696–700 (2002).
- Bosanac, I. *et al.* Crystal structure of the ligand binding suppressor domain of type 1 inositol 1,4,5-trisphosphate receptor. *Mol. Cell* **17**, 193–203 (2005).
- Rossi, A. M. *et al.* Synthetic partial agonists reveal key steps in IP₃ receptor activation. *Nature Chem. Biol.* **5**, 631–639 (2009).
- Uchida, K., Miyauchi, H., Furuichi, T., Michikawa, T. & Mikoshiba, K. Critical regions for activation gating of the inositol 1,4,5-trisphosphate receptor. *J. Biol. Chem.* **278**, 16551–16560 (2003).
- Schug, Z. T. & Joseph, S. K. The role of the S4–S5 linker and C-terminal tail in inositol 1,4,5-trisphosphate receptor function. *J. Biol. Chem.* **281**, 24431–24440 (2006).
- Chan, J. *et al.* Structural studies of inositol 1,4,5-trisphosphate receptor: coupling ligand binding to channel gating. *J. Biol. Chem.* **285**, 36092–36099 (2010).
- Yamazaki, H., Chan, J., Ikura, M., Michikawa, T. & Mikoshiba, K. Tyr-167/Trp-168 in type 1/3 inositol 1,4,5-trisphosphate receptor mediates functional coupling between ligand binding and channel opening. *J. Biol. Chem.* **285**, 36081–36091 (2010).
- Tung, C. C., Lobo, P. A., Kimlicka, L. & Van Petegem, F. The amino-terminal disease hotspot of ryanodine receptors forms a cytoplasmic vestibule. *Nature* **468**, 585–588 (2010).
- Ludtke, S. J. *et al.* Flexible architecture of IP₃R1 by cryo-EM. *Structure* **19**, 1192–1199 (2011).
- Amador, F. J. *et al.* Crystal structure of type I ryanodine receptor amino-terminal β -trefoil domain reveals a disease-associated mutation "hot spot" loop. *Proc. Natl Acad. Sci. USA* **106**, 11040–11044 (2009).

- Lobo, P. A. & Van Petegem, F. Crystal structures of the N-terminal domains of cardiac and skeletal muscle ryanodine receptors: insights into disease mutations. *Structure* **17**, 1505–1514 (2009).
- Taylor, C. W. & Tovey, S. C. IP₃ receptors: toward understanding their activation. *Cold Spring Harb. Perspect. Biol.* **2**, a004010 (2010).
- Anyatonwu, G. & Joseph, S. K. Surface accessibility and conformational changes in the N-terminal domain of type I inositol trisphosphate receptors: studies using cysteine substitution mutagenesis. *J. Biol. Chem.* **284**, 8093–8102 (2009).
- Tateishi, H. *et al.* Defective domain-domain interactions within the ryanodine receptor as a critical cause of diastolic Ca²⁺ leak in failing hearts. *Cardiovasc. Res.* **81**, 536–545 (2009).
- Sureshan, K. M. *et al.* Activation of IP₃ receptors by synthetic bisphosphate ligands. *Chem. Commun.* **14**, 1204–1206 (2009).
- Chan, J. *et al.* Ligand-induced conformational changes via flexible linkers in the amino-terminal region of the inositol 1,4,5-trisphosphate receptor. *J. Mol. Biol.* **373**, 1269–1280 (2007).
- Mayer, M. L. Glutamate receptors at atomic resolution. *Nature* **440**, 456–462 (2006).
- Lin, C. C., Baek, K. & Lu, Z. Apo and InsP₃-bound crystal structures of the ligand-binding domain of an InsP₃ receptor. *Nature Struct. Mol. Biol.* **18**, 1172–1174 (2011).
- Hamada, T., Bannister, M. L. & Ikemoto, N. Peptide probe study of the role of interaction between the cytoplasmic and transmembrane domains of the ryanodine receptor in the channel regulation mechanism. *Biochemistry* **46**, 4272–4279 (2007).
- Ramos-Franco, J., Galvan, D., Mignery, G. A. & Fill, M. Location of the permeation pathway in the recombinant type-1 inositol 1,4,5-trisphosphate receptor. *J. Gen. Physiol.* **114**, 243–250 (1999).
- Samsó, M., Feng, W., Pessah, I. N. & Allen, P. D. Coordinated movement of cytoplasmic and transmembrane domains of RyR1 upon gating. *PLoS Biol.* **7**, e85 (2009).
- Iwai, M., Michikawa, T., Bosanac, I., Ikura, M. & Mikoshiba, K. Molecular basis of the isoform-specific ligand-binding affinity of inositol 1,4,5-trisphosphate receptors. *J. Biol. Chem.* **282**, 12755–12764 (2007).
- Chu, A., Diaz-Munoz, M., Hawkes, M. J., Brush, K. & Hamilton, S. L. Ryanodine as a probe for the functional state of the skeletal muscle sarcoplasmic reticulum calcium release channel. *Mol. Pharmacol.* **37**, 735–741 (1990).
- Lai, F. A. & Meissner, G. The muscle ryanodine receptor and its intrinsic Ca²⁺ channel activity. *J. Bioenerg. Biomembr.* **21**, 227–246 (1989).
- Liu, Z. *et al.* Dynamic, inter-subunit interactions between the N-terminal and central mutation regions of cardiac ryanodine receptor. *J. Cell Sci.* **123**, 1775–1784 (2010).
- George, C. H. *et al.* Ryanodine receptor regulation by intramolecular interactions between cytoplasmic and transmembrane domains. *Mol. Biol. Cell* **15**, 2627–2638 (2004).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Allen and D. MacLennan for gifts of plasmids encoding RyR2 and RyR1, respectively. C.W.T. thanks T. Rahman and V. Konieczny for discussions. M.I. acknowledges K. Mikoshiba and T. Michikawa for long-standing support and discussions. This work was supported by grants from the Heart and Stroke Foundation of Ontario (T-7181) to M.I., National Institutes of Health Research (EY012347 and NS059969) to J.B.A., the Wellcome Trust (085295), the Biotechnology and Biological Sciences Research Council (BB/H009736) and the Medical Research Council (G0900049) to C.W.T. M.-D.S. is supported by postdoctoral fellowships from the Canadian Institutes of Health Research and the National Research Foundation of Korea (2009-352-E00006). A.M.R. is a fellow of Queens' College, Cambridge. M.I. holds a Canadian Research Chair in Cancer Structural Biology.

Author Contributions M.-D.S., N.I., P.B.S., M.I. and C.L. determined and analysed the structure of NT. S.V. prepared and characterized the full-length InsP₃R and chimaeras. A.M.R., S.A.K. and P.D. completed analyses of InsP₃ binding and related molecular biology. J.B.A., M.I. and C.W.T. supervised work in their respective laboratories, coordinated the project and, with input from other authors, wrote the paper.

Author Information The atomic coordinates for NT(Cys-less) of rat InsP₃R1 with and without InsP₃ bound have been deposited in the Protein Data Bank under accession numbers 3UJ4 and 3UJ0, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.W.T. (cwt1000@cam.ac.uk) or M.I. (mikura@uhnres.utoronto.ca).

METHODS

Materials. InsP₃ was from Enzo Life Sciences. Adenophostin A was from A. G. Scientific. Ryanodine was from Ascent Scientific. Cyclopiazonic acid was from Sigma. Sources of other materials are specified in earlier publications^{2–4,17} or described below.

Cloning, expression and purification of N-terminal fragments of InsP₃R1 and RyR2. The open reading frame (ORF) encoding the NT fragment (residues 1–604) of rat InsP₃R1 (GenBank accession number GQ233032.1) was amplified by polymerase chain reaction (PCR) from the full-length clone lacking the S1 splice site using forward 5'-CGGGATCCATGTCTGACAAAATGTCTAGT-3' and reverse 5'-CGCGCTCGAGTCACTTTCGGTTGTTGTGGA-3' primers. The PCR product was ligated into a pGEX-6P-2 vector (GE Healthcare) as a BamHI–XhoI fragment to give pGEX(NT), which includes an NT–GST tag followed by a PreScission-cleavage site. To generate NT(Cys-less), a QuikChange multisite directed mutagenesis kit (Agilent) was used to mutate all Cys residues to Ala using pGEX-6P-2(NT) as the template and using the primers that are listed in Supplementary Table 8. Residues are numbered by reference to rat InsP₃R1 containing the S1 splice site.

Plasmids encoding GST-tagged IBC (residues 224–604) and IBC(Cys-less) were generated using PCR to amplify the appropriate sequence from the ORF of full-length rat InsP₃R1 or NT(Cys-less) using the following primers: forward 5'-CGGGATCCATGAAATGGAGTACAAAG-3' and reverse 5'-CGCGCTCGAGTCACTTTCGGTTGTTGTGGA-3'. Each PCR product was ligated into a pGEX-6P-2 vector as a BamHI–XhoI fragment to produce pGEX-6P-2(IBC) and pGEX-6P-2(IBC(Cys-less)), respectively. For analysis of the effects of mutations within the SD on InsP₃ binding, the plasmids described previously were used to express His₆-tagged NT and IBC⁴ and the His₆-tag was cleaved before experiments⁴. Mutations were introduced using a QuikChange mutagenesis kit and using the primers listed previously⁴ or in Supplementary Table 9.

The sequence encoding the A domain of RyR2 was amplified by PCR from rabbit RyR2 (GenBank: GI164831)³⁰ in pcDNA3 using the primers forward 5'-ACTAGTCTCGAGGTGCTCTCCAGGGGCCCATGGCTGATGGGGGCGAA-3' and reverse 5'-GATATCCTTCACTTCTGAGCTGATGGG-3'. The ORF for the IBC of InsP₃R1 was excised from pGEX-6P-2(NT) as a BamHI–XhoI fragment and ligated into a pET41a vector to produce pET41a(IBC). To generate a plasmid encoding a chimaeric NT in which the A domain of RyR2 (residues 1–210) was fused to the IBC of InsP₃R1 (residues 225–604) (RyR2A-IBC), the PCR product from above was ligated into pET41a(IBC) as a SpeI–EcoRV fragment to produce pET41a(RyR2A-IBC). Mutations within the ORF of the A domain of RyR2A-IBC were generated by site-directed mutagenesis using the QuikChange Lightning mutagenesis kit (Stratagene) using the primers listed in Supplementary Table 9. The complete coding sequences of all constructs were confirmed by sequencing. The sequences of the proteins used are summarized in Supplementary Table 1 and Fig. 4a.

For structural studies, NT(Cys-less) was expressed as a GST-fusion protein in BL21-CodonPlus(DE3) *E. coli* strain. Transformed cells were first grown at 37 °C until the $D_{600\text{ nm}}$ reached ~1.0 and they were then induced with 0.5 mM IPTG at 15 °C for ~18 h. Proteins were purified using glutathione sepharose 4B resin (GE Healthcare), and the GST tag was cleaved from the eluted proteins with PreScission protease (GE Healthcare) during overnight dialysis at 4 °C in cutting buffer (20 mM Tris–HCl, pH 8.4, 300 mM NaCl, 5% glycerol and 2 mM DTT). The cleaved proteins were purified further with cation-exchange chromatography (Fractogel EMD SO3- resin, EM Industries) and size-exclusion chromatography (Superdex 200, GE Healthcare). Purified proteins were concentrated to 14 mg ml^{−1} in a buffer comprising 20 mM Tris–HCl, pH 8.4, 360 mM NaCl, 2.5% glycerol, 0.2 mM TCEP (tris(2-carboxyethyl)phosphine), 1 mM PMSF. Similar methods were used to express InsP₃R fragments for binding studies, but with modifications: bacteria were initially grown at 22 °C, the GST tag was cleaved by incubation of bacterial lysates that were immobilized on glutathione sepharose 4B resin with PreScission for 5 h in PreScission-cleavage buffer (GE Healthcare). The eluent was then used for ³H-InsP₃ binding analyses without further purification. Western blotting and silver-stained gels were used to verify expression and purification of NT fragments.

Western blot analysis. Western blotting of DT40 cells that were solubilized in Tris-EDTA medium (TEM) containing Triton X100 (1% v/v) was performed as previously described³¹ using anti-peptide antisera corresponding to residues 240–253 within the IBC (AbNT, 1:1,000) or 2733–2749 (AbCT, 1:500) of rat InsP₃R1. The secondary antibody was HRP-conjugated donkey anti-rabbit antibody (1:5,000, Santa Cruz Biotechnology).

³H-InsP₃ binding. Equilibrium-competition binding assays were performed at 4 °C in TEM (50 mM Tris, 1 mM EDTA, pH 8.3) containing ³H-InsP₃ (0.75 nM, Perkin-Elmer Life Sciences), purified protein (1–4 µg) and unlabelled InsP₃ in a final volume of 500 µl. After a 5-min incubation, during which equilibrium was

attained, reactions were terminated by addition of 500 µl of TEM containing 30% poly(ethylene glycol) 8000 and γ-globulin (750 µg), and centrifugation (20,000g, 5 min). For ³H-InsP₃ binding to IBC, the amount of ³H-InsP₃ was reduced to 0.25 nM, and incubation volumes were doubled. Pellets were solubilized in 200 µl of TEM containing 2% Triton X100 (v/v), mixed with EcoScintA scintillation liquid (National Diagnostics) and radioactivity was determined by liquid scintillation counting. Nonspecific binding was determined in the presence of 10 µM InsP₃. Binding results were fitted to a Hill equation (GraphPad Prism, version 5) from which pIC₅₀ (−logIC₅₀, in which IC₅₀ is the half-maximal inhibitory concentration) and thereby pK_d (−logK_d, where K_d is the equilibrium dissociation constant) values were calculated³².

³H-ryanodine binding. Microsomal membranes were prepared from DT40 cells by lysis with a glass homogenizer and sonication in cytosol-like medium (CLM) supplemented with protease inhibitors (Roche complete protease inhibitor cocktail), followed by centrifugation (50,000g, 30 min). The CLM contained 140 mM KCl, 20 mM NaCl, 1 mM EGTA, 20 mM PIPES, 2 mM MgCl₂, 375 µM CaCl₂ (free [Ca²⁺] of ~220 nM), pH 7. Equilibrium-competition binding was performed with microsomal membranes (100 µg protein per ml) at 4 °C in 200 µl of CLM supplemented with protease inhibitors and ³H-ryanodine (100 nM, Perkin-Elmer Life Sciences). Reactions were terminated after 90 min, and radioactivity was determined as described for ³H-InsP₃ binding. Nonspecific binding was defined by addition of 10 µM unlabelled ryanodine.

³H-ryanodine binding to RyR typically requires many hours to reach equilibrium²⁴ because it binds only to the open state of the channel and spontaneous openings are rare. In our analyses of ³H-ryanodine binding to InsP₃R1–RyR1 (Fig. 4g), equilibrium was attained within 90 min, perhaps because the modestly increased spontaneous activity of the chimaeric channel (Supplementary Fig. 10b) contributed to an increased rate of ³H-ryanodine binding to the open state. In parallel comparisons, specific binding of ³H-ryanodine to InsP₃R1–RyR1 expressed in DT40 cells and stimulated with InsP₃ (1 µM) was (d.p.m.: mean (range) for 2 independent experiments): 4,241 (4,073–4,409) after 90 min, 4,941 (4,825–5,058) after 3 h, and 4,410 (4,108–4,712) after 14 h. It is, however, important to note that our conclusion that InsP₃ selectively stimulates ³H-ryanodine binding to InsP₃R1–RyR1 is not dependent on having measured binding under equilibrium conditions.

Crystallization and data collection. Crystals of apo-NT(Cys-less) were grown by the hanging-drop vapour diffusion method at 293 K by mixing 1 µl of protein with an equal volume of reservoir solution (0.1 M Hepes, pH 7.0, 0.8–1.0 M (NH₄)₂SO₄). Using an additives screen, 3% (v/v) trimethylamine N-oxide was identified as an important additive to obtain single rod-shaped crystals. After a series of microseeding trials, rod-shaped single crystals were obtained within 5 days. For crystallization of InsP₃-bound NT(Cys-less), five molar excess of InsP₃ (~1 mM) was added before crystallization. Crystals of InsP₃-bound NT(Cys-less) were grown using the same method except for the reservoir solution containing 0.1 M Na citrate (pH 6.0), 8% (w/v) PEG-6000, 70 mM Li₂SO₄ and 3% dimethyl sulphoxide.

For data collection, crystals were equilibrated in 25% glycerol cryoprotective solutions containing reservoir buffer, and flash frozen in liquid nitrogen. Diffraction data were collected at 100 K on a 19-ID beamline for apo-state crystals or 19-BM beamline for InsP₃-bound crystals at the Advanced Photon Source Synchrotron facility (Argonne National Laboratory), and were processed with HKL2000. Crystals of apo-NT(Cys-less) belong to the space group P1 with cell dimensions $a = 63.1 \text{ \AA}$, $b = 77.2 \text{ \AA}$, $c = 101.5 \text{ \AA}$, $\alpha = 105.4^\circ$, $\beta = 100.0^\circ$, $\gamma = 101.0^\circ$. Crystals of InsP₃-bound NT(Cys-less) belong to the space group C2 with cell dimensions $a = 189.2 \text{ \AA}$, $b = 78.7 \text{ \AA}$, $c = 134.1 \text{ \AA}$, $\alpha = 90.0^\circ$, $\beta = 124.5^\circ$, $\gamma = 90.0^\circ$. Crystals of both apo- and InsP₃-bound NT(Cys-less) contained two molecules in the asymmetric unit (Supplementary Table 3).

Structure determination and refinement. Structures of apo-NT(Cys-less) at 3.0 Å resolution and of InsP₃-bound NT(Cys-less) at 3.6 Å resolution were determined by molecular replacement using structures of the SD (PDB code 1XZZ)³ and the IBC (PDB code 1N4K)² as search models with the program Phaser²⁹. Iterative refinement and model building were performed with Refmac5 (ref. 33) and Coot³⁴, respectively (Supplementary Table 3). Structures of the two molecules in the asymmetric unit of apo-NT(Cys-less) are virtually identical (root mean squared deviation (r.m.s.d.) value = 0.543 Å) except for a minor variation in the loop between β-strand 20 and β-strand 21 that does not affect the interpretation of our results. The low r.m.s.d. between chain A and chain B is maintained through the regions of the molecule that make up the α-interface and β-interface, thus increasing the validity of our description of the 'open-clam' structure. The two molecules in the asymmetric unit of InsP₃-bound NT(Cys-less) are more converged than those of the apo-structure (r.m.s.d. value = 0.134 Å), which also validates our description of the 'closed-clam' structure. The molecule of chain A for each state was used to generate figures, but the chain B molecule of apo-NT(Cys-less) was used for the side chain of Asp444 in Fig. 1c. All water molecules

were modelled in Coot³⁴. Initially, water molecules were detected using the automatic 'find waters' function in the program. A $2F_o - F_c$ map was used with a sigma cutoff value of 1.0, and minimum and maximum distances to protein atoms of 2.4 Å and 3.2 Å, respectively. We subsequently picked additional water molecules and deleted inappropriate water molecules by manually surveying the density in Coot. After refinement, all water molecules exhibiting negative electron density due to inconsistent modelling were deleted.

Circular dichroism analysis. Circular dichroism spectra were collected on a Jasco J-720 spectrometer using a 1-mm path length cuvette at 20 °C. The NT and NT(Cys-less) (0.2 mg ml⁻¹) were prepared in a buffer (20 mM Tris-HCl, pH 8.4, 360 mM NaCl, 2.5% glycerol, 0.2 mM TCEP, 1 mM PMSF). Circular dichroism spectra were obtained from 260 to 200 nm, with a 2-nm bandwidth, an 8-s response time and a scan speed of 50 nm min⁻¹. Data are averages of three consecutive scans.

Cloning and functional expression of chimaeric InsP₃R. To generate the plasmid encoding a chimaeric InsP₃R1 in which residues 2274–2748 of InsP₃R1 (all residues downstream of those immediately before TMD1) were replaced by the equivalent region from RyR1 (residues 4511–5037) (InsP₃R1-RyR1), the appropriate region of the ORF of rabbit RyR1 (GenBank accession number X15209)³⁵ was amplified by PCR from the expression vector pcDNA3.2 using the primers forward, 5'-CGCGGGTTCGAAGTCCCCGAGGCCACCAGAACCCCC-3', and reverse 5'-CGGGGCGTCCTCGAGTCATTAGCTCAGCTGGTCTCGTACTGCTTGCGGAAGC-3'. The PCR product was cloned in-frame as a BstBI-XhoI fragment into a pENTR1a vector containing nucleotides 1–6822 of rat InsP₃R1. This construct was transferred into the Gateway-compatible expression vector, pcDNA3.2, to generate pcDNA3.2-(InsP₃R1-RyR1). A plasmid encoding InsP₃R that lacked the SD was generated from ORFs for the full-length InsP₃R1 lacking the S1 splice site (pENTR1A(InsP₃R1)) and the IBC (pENTR1A(IBC)). Both plasmids were digested with NheI and KpnI, and the fragment from pENTR1A(IBC) was cloned into pENTR1A(InsP₃R1). Site-directed mutagenesis was then used to silence 3 internal BamHI sites within this construct without affecting the coding sequence to generate the plasmid pENTR1A(InsP₃R1^{ΔSD}). A plasmid encoding a chimaera in which the SD of InsP₃R1 (residues 1–224) was replaced by the A domain of RyR1 (residues 1–210) (RyR1A-InsP₃R) was prepared by isolating the coding sequence for the A domain of RyR1 by PCR from the rabbit RyR1 ORF using the primers forward, 5'-GCTAGCATCATGGGTGACGGAGGA-3' and reverse 5'-GGATCCTTCACAGCAGGAGCAGATG-3'. The PCR product was cloned as a NheI-BamHI fragment into pENTR1A(InsP₃R1^{ΔSD}). The complete coding sequences of all plasmids were verified by sequencing. Domain boundaries of the chimaeric proteins are summarized in Supplementary Table 1.

DT40 cells lacking functional genes for native InsP₃Rs (DT40 knockout (DT40-KO)) were transfected by electroporation with linearized plasmids (10 µg DNA per 10⁶ cells) using the Neon (Invitrogen) or Amaxa (Lonza) nucleofection systems. G418 (2 mg ml⁻¹) was used to select and amplify clones of G418-resistant cells. Stable cell lines were selected and InsP₃R expression was measured by western blotting. DT40 cells were cultured in RPMI 1640 medium with L-glutamine (Invitrogen) supplemented with 10% fetal bovine serum, 1% heat-inactivated chicken serum (both from Sigma) and 10 µM 2-mercaptoethanol at 37 °C in humidified air containing 5% CO₂. Cells (~2 × 10⁶ cells per ml) were passaged

every 2–3 days. Similar methods were used for transient transfections with RyR1 and InsP₃R1-NT(Cys-less), but with 50 µg DNA per 3 × 10⁶ cells.

Functional analyses of InsP₃R in DT40 cells. Uptake of Ca²⁺ into the intracellular stores of saponin-permeabilized DT40 cells and its release by InsP₃ were measured using a low-affinity Ca²⁺ indicator (Mag-fluo-4) trapped within the endoplasmic reticulum³⁶. All experiments were performed at 20 °C in CLM supplemented with 1.5 mM MgATP to allow active Ca²⁺ uptake. After the intracellular stores had loaded to steady state with Ca²⁺ (~150 s), InsP₃ was added with thapsigargin (1 µM) to prevent further Ca²⁺ uptake (Supplementary Fig. 9). The effects of InsP₃ were assessed after a further 10–40 s. InsP₃-evoked Ca²⁺ release is expressed as a fraction of the ATP-dependent Ca²⁺ uptake. Typical experiments are shown in Supplementary Fig. 9.

Structural model of RyR2A-IBC. A structural homology model of RyR2A-IBC (Fig. 4c) was produced using UCSF Chimera³⁷ to first superpose the backbone structures of apo-NT(Cys-less) and the ABC of RyR1 (PDB, 2XOA), the only RyR subtype for which there is a complete NT structure⁹. This A-domain structure of RyR1 was then used to allow superposition of the A domain from RyR2 (PDB, 3IM5)¹², effectively achieving superposition of NT(Cys-less) onto a 'virtual' chimaera of RyR2A with RyR1BC. The predicted structure of the RyR2A-IBC was then revealed by masking the SD of InsP₃R1 and the BC domains of RyR1 (Fig. 4c).

Computational docking. Rigid-body docking of the apo-NT(Cys-less) structure into a ~10 Å cryo-electron microscopy density map of InsP₃R1 (EMDB, EMD-5278) was implemented using the six-dimensional search procedure in the Situs Program package³⁸. The Laplacian filter was applied to the density maps to enhance the fitting contrast. Docking of the crystal structure of RyR1-ABC (PDB, 2XOA) into a 9.6 Å cryo-electron microscopy density map (EMDB, EMD-1275), as previously described⁹, was repeated using the above procedure. The UCSF Chimera package³⁷ was used to visualize the docking results with the density maps (Fig. 3 and Supplementary Fig. 6).

- Otsu, K. *et al.* Molecular cloning of cDNA encoding the Ca²⁺ release channel (ryanodine receptor) of rabbit cardiac muscle sarcoplasmic reticulum. *J. Biol. Chem.* **265**, 13472–13483 (1990).
- Tovey, S. C. *et al.* Regulation of inositol 1,4,5-trisphosphate receptors by cAMP independent of cAMP-dependent protein kinase. *J. Biol. Chem.* **285**, 12979–12989 (2010).
- Kenakin, T. P. *Pharmacologic Analysis of Drug-Receptor Interactions* 3rd edn (Lippincott, Williams & Wilkins, 1997).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Zorzato, F. *et al.* Molecular cloning of cDNA encoding human and rabbit forms of the Ca²⁺ release channel (ryanodine receptor) of skeletal muscle sarcoplasmic reticulum. *J. Biol. Chem.* **265**, 2244–2256 (1990).
- Tovey, S. C., Sun, Y. & Taylor, C. W. Rapid functional assays of intracellular Ca²⁺ channels. *Nature Protocols* **1**, 259–263 (2006).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).

Biosignatures as revealed by spectropolarimetry of Earthshine

Michael F. Sterzik¹, Stefano Bagnulo² & Enric Pallé³

Low-resolution intensity spectra of Earth's atmosphere obtained from space reveal strong signatures of life ('biosignatures'), such as molecular oxygen and methane with abundances far from chemical equilibrium, as well as the presence of a 'red edge' (a sharp increase of albedo for wavelengths longer than 700 nm) caused by surface vegetation¹. Light passing through the atmosphere is strongly linearly polarized by scattering (from air molecules, aerosols and cloud particles) and by reflection (from oceans and land²). Spectropolarimetric observations of local patches of Earth's sky light from the ground contain signatures of oxygen, ozone and water, and are used to characterize the properties of clouds and aerosols^{3,4}. When applied to exoplanets, ground-based spectropolarimetry can better constrain properties of atmospheres and surfaces than can standard intensity spectroscopy⁵⁻⁹. Here we report disk-integrated linear polarization spectra of Earthshine, which is sunlight that has been first reflected by Earth and then reflected back to Earth by the Moon¹⁰⁻¹³. The observations allow us to determine the fractional contribution of clouds and ocean surface, and are sensitive to visible areas of vegetation as small as 10 per cent. They represent a benchmark for the diagnostics of the atmospheric composition, mean cloud height and surfaces of exoplanets.

Broadband polarimetric measurements of the disk-integrated Earth have been obtained previously¹⁴. Observations of Earthshine at a wavelength of 550 nm reported a fraction of linear polarization of about 10% for a Sun–Earth–Moon phase angle near to 90° (quadrature). As the lunar surface is responsible for a depolarization of the incident light by a factor of ~3.3, the true linear polarization of the disk-integrated light of Earth was estimated to be around 30–35% (ref. 14). Modern estimates of the Earth's linear polarization are based on measurements by the satellite-borne instrument POLDER in three bands (443 nm, 670 nm and 865 nm), and demonstrate strong dependence on cloud cover and on wavelength¹⁵. These measurements, however, do not allow a spectral analysis, and therefore cannot be used to infer biosignatures.



Using the Focal Reducer/Low-dispersion Spectrograph (FORS)¹⁶ mounted at the Very Large Telescope in Chile, we measured the linear polarization spectra of Earthshine. Our data were collected during two observing epochs with distinct viewing geometries of Earth (observing techniques and data reduction details are given in Supplementary Information). A detailed breakdown of the surfaces visible in our Earthshine observations, derived from MODIS satellite observations¹⁷, is given in Table 1.

The red curves in Figs 1 and 2 display the wavelength (λ) dependence of the fraction of linear polarization, $P_Q(\lambda)$ ($=Q/I$, the polarized flux Q over the total flux I , see the definition of the Stokes parameters in Supplementary Information), measured in April (Fig. 1) and in June (Fig. 2). The shape of the $P_Q(\lambda)$ profile depends mainly on Earth's surface albedo at the time of observations. On 25 April 2011, the continuum polarization reaches about 9% at 500 nm, and decreases smoothly to 4% at 900 nm. The polarization observed on 10 June 2011 is 3% higher than observed in April. A change of the continuum slope occurs visibly in the June spectrum at a wavelength of around 520 nm.

In particular, the June spectrum matches the observation by Dollfus¹⁴ of 10% Earthshine polarization at his observing wavelength of 550 nm, whereas our April value is smaller. In order to display the detailed structure on top of the smooth continuum, we have fitted and subtracted from the observed $P_Q(\lambda)$ profile a fourth-order polynomial (between 530 and 910 nm). The residual signal $\delta P(\lambda)$ shows a rich structure at the 0.1% level. A narrow O₂ A-band (760 nm) is prominent in both spectra at a level of almost 1% above the continuum. O₂ B (690 nm) and H₂O (720 and 820 nm) bands are broader and with lower amplitude. An additional (strong) polarization feature is present in the June spectrum close to 850 nm, and a (weak) feature at 854 nm. We tentatively attribute them to the presence of a variable, thin Ca II layer in Earth's ionosphere, recently discovered through observations of the transmitted spectrum of the Earth¹⁸. Remarkably, the overall shape of the residuals between 550 and 750 nm differs between the two epochs: whereas the April spectrum exhibits a broad, shallow peak around 700 nm, the June spectrum appears rather flat.

We compared our measurements with the polarization estimated from POLDER data¹⁵, and with published tables of radiative transfer models of Earth-like extrasolar planets⁵. For comparison purposes, both POLDER data and model predictions have to be scaled to account for the depolarization of the Earthshine caused by reflection from the

Table 1 | Earth observations

Observations	Observing date	
	25 April 2011, 09:00 UT	10 June 2011, 01:00 UT
View of Earth as seen from the Moon		
Sun–Earth–Moon phase (degrees)	87	102
Ocean fraction in Earthshine (%)	18	46
Vegetation fraction in Earthshine (%)	7	3
Tundra, shrub, ice and desert fraction in Earthshine (%)	3	1
Total cloud fraction in Earthshine (%)	72	50
Cloud fraction $\tau > 6$ (%)	42	27

This Table gives the contribution of different types of surface to the Earthshine, determined by mapping satellite cloud retrievals from MODIS on the parts of the Earth's surface that are visible from the Moon at the given observing epoch¹⁷. The first observing epoch was before dawn on 25 April 2011, and the Earthshine contains contributions from the Atlantic Ocean, the Amazonas region, and parts of Europe and Africa. The second observing epoch was after dusk on 10 June 2011, and probes the Pacific side of the Earth, with almost no visible land surfaces. The contribution to the Earthshine signal of a given Earth region will depend on the solar and lunar zenith angles (the reflectance is a bi-directional property), thus the percentage scene types and cloud cover fractions in the Table have been averaged over the whole Earthshine-contributing area using these weights¹³. The contributions of clear tundra, shrub, desert or ice scenes were negligible.

¹European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago, Chile. ²Armagh Observatory, College Hill, Armagh BT61 9DG, UK. ³Instituto de Astrofísica de Canarias, Vía Lactea s/n, E38205 La Laguna, Tenerife, Spain.

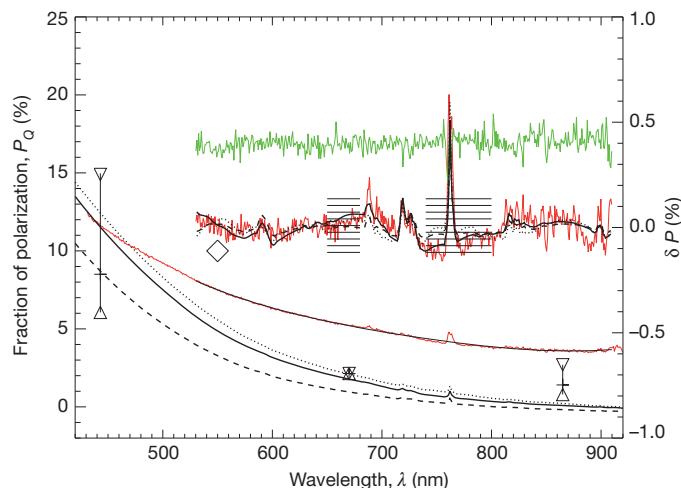


Figure 1 | Polarization spectra observed on 25 April 2011. Red lines show observed fractional polarization P_Q , black lines show model spectra for comparison. Both continuum polarization and residual polarization (derived from the continuum by subtracting a polynomial fitted in the wavelength region between 530 and 910 nm) are plotted with different scale units on the left and right axis to zoom in on the spectral fine-structure. For direct comparison with the Earthshine, the model spectra have been reduced in amplitude to account for the depolarization caused by the lunar surface^{19,20}; the different linestyles refer to different choices of the model parameters (Table 2). The green line refers to the measured Stokes P_U as explained in the Supplementary Information (same scale as the residuals on the right axis), and is representative for the noise level in the spectra. Horizontal hatching between 650–680 nm and 740–800 nm denotes the NDVI bandpass regions and refers to the corresponding residual polarizations. Open triangles, POLDER-based whole-Earth polarization estimates at 443 nm, 670 nm and 865 nm with a mean cloud coverage of 55%; upper and lower triangles bracket low (10%) or high (90%) cloud coverage levels. Open diamond, Dollfus's estimate of an Earthshine polarization of 10% at a phase angle of 80° at 550 nm. An NDVI contrast is visible in the data and for model 'a', which is consistent with a 10–15% visible vegetation content in the Earthshine.

lunar surface. The lunar depolarization factor is not well determined, but roughly correlates with the albedo of the reflecting surfaces^{19,20}. Following a classic study¹⁴, we assume a linear increase of depolarization (depol) with wavelength, normalized to 3.3 at a wavelength at 550 nm (depol = $3.3\lambda/550$, where λ is in units of nm).

The interpretation of the POLDER data depends on the cloud coverage at the time of observation and is roughly consistent with our continuum values in April, but too low in June. The higher

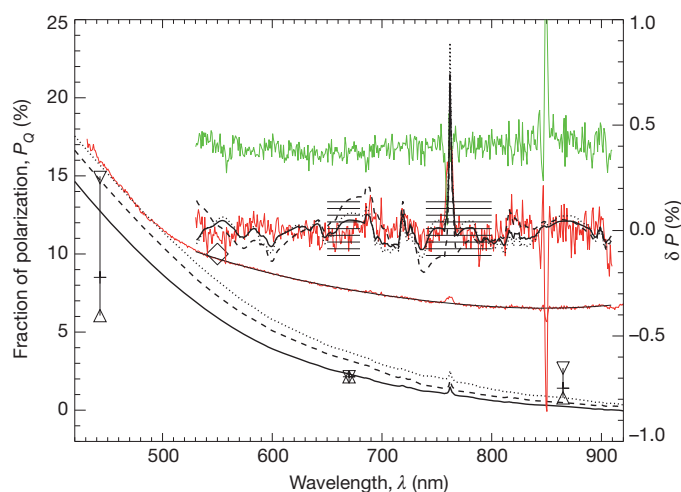


Figure 2 | Polarization spectra observed on 10 June 2011. Lines and symbols as for Fig. 1. An NDVI contrast is not evident in the data, and models 'a' and 'c' are consistent with no visible vegetation content in the Earthshine.

polarization we observed in June can be qualitatively explained by an approximately 20% lower cloud coverage during our Earthshine measurements than assumed for the disk-integration of the POLDER data¹⁵.

For the comparisons with theoretical predictions, we constructed a fine grid of possible Earth spectra from published tables by varying the parameters in steps of 1% for the different scene types⁵ (ocean clouded or clear, and vegetation clouded or clear). Then we compared these models with our observations in the wavelength region between 420 nm and 530 nm. In the blue spectral range, the effect of cloud properties on polarization is minor²¹, and synthetic Earth model polarized spectra agree quantitatively with our observations. The parameters of the best fitting models are listed in Table 2 (April c and June c). For both observing epochs, the modelling results match MODIS-derived cloud coverage fractions calculated for clouds having optical depth $\tau > 6$. The June model—less affected by clouds—approximates the continuum well. However, the models⁵ systematically predict polarizations in the red part of the continuum that are too low, primarily because of the approximations used in the models for the cloud microphysics²². Real clouds are patchy and, in particular over the ocean, exhibit much larger mean effective radii (droplet sizes 12–15 μm) and a high variation in their optical depth distribution²³. Depending on the underlying surface types, higher polarization fractions are thus expected.

Guided by the actual observed surface scene fractions (Table 1), we have calculated more polarization spectra from tabulated model spectra⁵ with parameters that are other plausible approximations to the real surface fractions responsible for the Earthshine spectrum at the epoch of the observations (model class parameters 'a' and 'b'). Although these models fail to explain the continuum, their residual profiles (generated with the same fourth-order polynomial fit and subtraction from the continuum as adopted as for the observed spectra) are consistent with our observations.

In Figs 1 and 2, we have hatched (horizontal lines) two bandpass regions at 650–680 nm and 740–800 nm and calculated the normalized difference vegetation index (NDVI)²⁴. A larger contrast in the continuum averages in April than in June is interpreted by the presence of a strong vegetation 'red edge' signal in April, which is practically undetected in June. This feature in the April spectra can be reproduced by assuming about 10–15% cloud-free land vegetation (model April a of Table 2). A smaller fraction (models April b and c) reduces the NDVI contrast too much, while too large a fraction would overestimate the observed feature. Similarly, the June data can best be described through little or absent cloud-free vegetation surfaces (models June a and c), while a 10% cloud-free vegetation fraction produces a NDVI that is too pronounced to be compatible with the data (model June b). We estimate that our data are sensitive to variations of 5% in the NDVI. Also, the observed polarization peak of the O_2 A feature (and to a lesser extent, O_2 B) is fully compatible with the models. The amplitude of these features are sensitive to the top of the cloud layer, and to the O_2 mixing ratio. Our Earthshine observations suggest a low top-layer (around 800 hPa), assuming the mixing ratio on Earth to be fixed⁵. The H_2O and O_3 (580 nm) band strengths also appear fully consistent with the model.

Table 2 | Model parameters

Model	Ocean clouded (%)	Ocean clear (%)	Vegetation clouded (%)	Vegetation clear (%)	Line in Fig. 2
April a	48	40	0	12	Solid
April b	60	30	10	0	Dashed
April c	44	56	0	0	Dotted
June a	40	60	0	0	Solid
June b	30	60	0	10	Dashed
June c	27	73	0	0	Dotted

This Table gives fractions of different types of surfaces used to model polarization spectra published in tables in ref. 5. Parameters for 'a' and 'b' models (leftmost column) are guided by observed scene type fractions in Table 1. The 'c' models are generated through a grid of synthetic spectra and fit the observations best within 420 and 530 nm. Note that the ocean clouded fractions agree well with the observed cloud fractions for optical depths $\tau > 6$ in Table 1.

Our spectropolarimetric measurements of the Earthshine constrain the Earth's surface and atmospheric composition and biosignatures. Improved vector radiative transfer models with more realistic cloud and surface treatment are necessary to fully account for the observed spectra. Validation of these models is necessary to interpret Earth-like exoplanet polarization spectra.

Received 16 September; accepted 9 December 2011.

- Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D., & Hord, C. A search for life on Earth from the Galileo spacecraft. *Nature* **365**, 715–721 (1993).
- Coffeen, D. L. Polarization and scattering characteristics in the atmospheres of Earth, Venus, and Jupiter. *J. Opt. Soc. Am.* **69**, 1051–1064 (1979).
- Aben, I., Helder, F., Stam, D. M. & Stammes, P. Spectral fine-structure in the polarization of skylight. *Geophys. Res. Lett.* **26**, 591–594 (1999).
- Boesche, E. *et al.* Polarization of skylight in the O₂A band: effects of aerosol properties. *Appl. Opt.* **47**, 3467–3480 (2008).
- Stam, D. M. Spectropolarimetric signatures of Earth-like extrasolar planets. *Astron. Astrophys.* **482**, 989–1007 (2008).
- Buenzli, E. & Schmid, H. M. A grid of polarization models for Rayleigh scattering planetary atmospheres. *Astron. Astrophys.* **504**, 259–276 (2009).
- Keller, C. U. *et al.* EPOL: the exoplanet polarimeter for EPICS at the E-ELT. *Proc. SPIE* **7735**, 77356G (2010).
- Bailey, J. Rainbows, polarization, and the search for habitable planets. *Astrobiology* **7**, 320–332 (2007).
- Williams, D. M. & Gaidos, E. Detecting the glint of starlight on the oceans of distant planets. *Icarus* **195**, 927–937 (2008).
- Arnold, L., Gillet, S., Lardiere, O., Riaud, P. & Schneider, J. A test for the search for life on extrasolar planets. Looking for the terrestrial vegetation signature in the Earthshine spectrum. *Astron. Astrophys.* **392**, 231–237 (2002).
- Woolf, N. J., Smith, P. S., Traub, W. A. & Jucks, K. W. The spectrum of Earthshine: a pale blue dot observed from the ground. *Astrophys. J.* **574**, 430–433 (2002).
- Seager, S., Turner, E. L., Schafer, J. & Ford, E. B. Vegetation's red edge: a possible spectroscopic biosignature of extraterrestrial plants. *Astrobiology* **5**, 372–390 (2005).
- Montanes-Rodriguez, P., Palle, E., Goode, P. R. & Martin-Torres, F. J. Vegetation signature in the observed globally integrated spectrum of Earth considering simultaneous cloud data: applications for extrasolar planets. *Astrophys. J.* **651**, 544–552 (2006).
- Dollfus, A. Études des planètes par la polarisation de leur lumière. *Suppléments aux Annales d'Astrophysique* **4**, 3–114 (1957).
- Wolstencroft, R. D. & Breon, F.-M. in *Astronomical Polarimetry: Current Status and Future Directions* (eds Adamson, A., Aspin, C., Davis, C. J. & Fujiyoshi, T.) 211–212 (ASP Conf. Ser. Vol. 343, Astronomical Society of the Pacific, 2005).
- Appenzeller, I. *et al.* Successful commissioning of FORS1 — the first optical instrument on the VLT. *ESO Messenger* **94**, 1–6 (1998).
- Giovanni portal. <http://disc.sci.gsfc.nasa.gov/giovanni/overview/index.html>.
- Pallé, E., Zapatero Osorio, M. R., Barrena, R., Montanes-Rodriguez, P. & Martin, E. L. Earth's transmission spectrum from lunar eclipse observations. *Nature* **459**, 814–816 (2009).
- Fox, G. K. *et al.* Solar system observations by the Wisconsin Ultraviolet Photopolarimeter Experiment — III. The first ultraviolet linear spectropolarimetry of the Moon. *Mon. Not. R. Astron. Soc.* **298**, 303–309 (1998).
- Shkuratov, Y. *et al.* Optical measurements of the Moon as a tool to study its surface. *Planet. Space Sci.* **59**, 1326–1371 (2011).
- Goloub, P., Deuze, J. L., Herman, M. & Fouquart, Y. Analysis of the POLDER polarization measurements performed over cloud covers. *IEEE Trans. Geosci. Rem. Sens.* **32**, 78–88 (1994).
- Karalidi, T., Stam, D. M. & Hovenier, J. W. Flux and polarisation spectra of water clouds on exoplanets. *Astron. Astrophys.* **530**, A69 (2011).
- Kokhanovsky, A., Platnick, S. & King, M. D. in *The Remote Sensing of Tropospheric Composition from Space* (eds Burrows, J. P. *et al.*) 231–257 (Springer, 2011).
- Tinetti, G. *et al.* Detectability of planetary characteristics in disk-averaged spectra II: synthetic spectra and light-curves of Earth. *Astrobiology* **6**, 881–900 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This Letter is based on data collected with the Very Large Telescope under ESO programme 87.C-0040(A) and (B). E.P. acknowledges support from the Spanish MICIIN, grant CGL2009-10641.

Author Contributions M.F.S. was the principal investigator of the programme, contributing to observation preparation and execution, data analysis and interpretation, comparison with models, and writing of the manuscript. S.B. was a co-investigator, contributing to observation preparation and execution, data reduction and interpretation, error analysis, comparison with models, and also contributing to the manuscript. E.P. was a co-investigator, contributing to observation preparation, data interpretation, and derivations of actual Earth surface fractions during the observations.

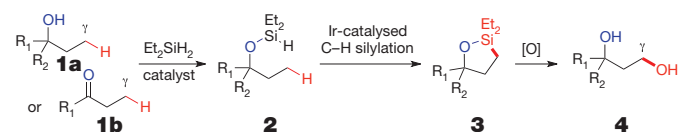
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.F.S. (msterzik@eso.org).

Catalytic functionalization of unactivated primary C–H bonds directed by an alcohol

Eric M. Simmons^{1†} & John F. Hartwig^{1†}

New synthetic methods for the catalytic functionalization of C–H bonds have the potential to revolutionize the synthesis of complex molecules^{1–4}. However, the realization of this synthetic potential requires the ability to functionalize selectively one C–H bond in a compound containing many such bonds and an array of functional groups. The site-selective functionalization of aliphatic C–H bonds is one of the greatest challenges that must be met for C–H bond functionalization to be used widely in complex-molecule synthesis^{1,3,5,6}, and processes catalysed by transition-metals provide the opportunity to control selectivity^{7,8}. Current methods for catalytic, aliphatic C–H bond functionalization typically rely on the presence of one inherently reactive C–H bond^{9,10}, or on installation and subsequent removal of directing groups that are not components of the desired molecule⁸. To overcome these limitations, we sought catalysts and reagents that would facilitate aliphatic C–H bond functionalization at a single site, with chemoselectivity derived from the properties of the catalyst and site-selectivity directed by common functional groups¹¹ contained in both the reactant and the desired product. Here we show that the combination of an iridium-phenanthroline catalyst and a dihydrosilane reagent leads to the site-selective γ -functionalization of primary C–H bonds controlled by a hydroxyl group, the most common functional group in natural products¹². The scope of the reaction encompasses alcohols and ketones bearing many substitution patterns and auxiliary functional groups; this broad scope suggests that this methodology will be suitable for the site-selective and diastereoselective functionalization of complex natural products.

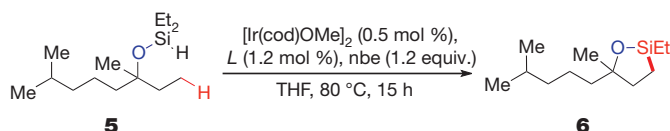
Our strategy for the functionalization of aliphatic C–H bonds directed by hydroxyl groups is outlined below:



In this approach, a single dihydrosilane reagent has two roles. First, the silane attaches to the oxygen atom of an alcohol (**1a**) or ketone (**1b**) by forming a (hydrido)silyl ether (**2**) to direct the C–H bond functionalization. The (hydrido)silyl ether is formed by dehydrogenative coupling with the alcohol or by hydrosilylation of the ketone. Second, the Si–H unit of the silyl ether undergoes dehydrogenative functionalization of a primary C–H bond without isolation of the intermediate (hydrido)silyl ether. Tamao–Fleming oxidation of the oxasilolane (**3**) formed by this process then yields a 1,3-diol (**4**) from a net 1,4-hydroxyl-directed, aliphatic C–H bond functionalization. On the basis of our recently developed, iridium-catalysed silylation of aromatic C–H bonds¹³, we anticipated that both silylation reactions could be catalysed by iridium complexes^{13,14}. However, successful implementation of this strategy would require the identification of a catalyst that is reactive for the silylation of unactivated aliphatic C–H bonds, not just the silylation of typically more reactive aromatic C–H bonds. It would also require that the directing effect be strong enough to ensure that intermolecular

functionalization of aromatic C–H bonds would not interfere with the desired directed functionalization of an aliphatic C–H bond.

We began by evaluating iridium complexes containing a series of bipyridine ligands (*L*) as catalysts for the dehydrogenative cyclization of (hydrido)silyl ether (**5**), generated *in situ* by an iridium-catalysed dehydrogenative coupling of diethylsilane (Et_2SiH_2) with tetrahydro-1,10-linalool. The reaction is outlined below:



In the presence of 1.0 mol % of a combination of $[\text{Ir}(\text{cod})\text{OMe}]_2$ (cod, 1,5-cyclooctadiene; Me, methyl group) and 1,10-phenanthroline (phen), with norbornene (nbe) as an H_2 acceptor¹³, we observed 77% conversion of **5** to form oxasilolane (**6**) in 68% yield (Table 1a). The same reaction conducted with 4,4'-di-*tert*-butylbipyridine (dtbpy), the most commonly used ligand for iridium-catalysed arene borylation¹⁵, occurred with only slightly higher conversion and yield (Table 1b). To probe systematically the effect of the electronic properties of ancillary ligands on iridium-catalysed aliphatic C–H bond silylation, we conducted reactions with catalysts generated from a series of 4,7-disubstituted phenanthroline derivatives (Table 1c–g). Reactions conducted with the more electron-donating ligands occurred to higher relative conversions and formed the silylation product in higher yields than did reactions conducted with the less electron-donating ligands¹⁶. Ultimately, we obtained the highest yield of **6** with a catalyst containing 3,4,7,8-tetramethyl-1,10-phenanthroline (Me_4phen , Table 1h), the most electron-donating phenanthroline derivative in this series¹⁶.

After identifying a suitable catalyst for aliphatic C–H bond silylation, we developed a telescoped procedure for transformation of a starting alcohol (**7**) or ketone (**8**) to a 1,3-diol (Fig. 1). Following iridium-catalysed aliphatic C–H silylation, we obtained the diol simply by adding MeOH, KHCO_3 and aqueous H_2O_2 to the crude silylation reaction mixture and heating at 50 °C. To provide compounds that were easily purified by silica-gel chromatography, we acylated the diols thus obtained before isolation. Figure 1 shows a series of products

Table 1 | Examination of bipyridine ligands for aliphatic C–H bond silylation (5** → **6**)**

	Ligand, <i>L</i>	Conversion	Yield
a	phen	77%	68%
b	dtbpy	83%	76%
c	4,7- Cl_2 phen	13%	<2%
d	4,7-(HO) ₂ phen	33%	19%
e	4,7- Ph_2 phen	82%	72%
f	4,7- Me_2 phen	94%	85%
g	4,7-(MeO) ₂ phen	94%	87%
h	3,4,7,8- Me_4 phen	100%	99%

Corrected, overall gas chromatography yields (0.25-mmol scale) with dodecane as an internal standard. Reagents and conditions: tetrahydro-1,10-linalool (1.0 equiv.), Et_2SiH_2 (1.2 equiv.), $[\text{Ir}(\text{cod})\text{OMe}]_2$ (0.05 mol %), THF, room temperature, 17 h; removal of volatiles, then $[\text{Ir}(\text{cod})\text{OMe}]_2$ (0.5 mol %), ligand (1.2 mol %), nbe (1.2 equiv.), THF, 80 °C, 15 h. For full experimental details, see the Supplementary Information.

¹Department of Chemistry, University of Illinois, Urbana, Illinois 61801, USA. [†]Present address: Department of Chemistry, University of California, Berkeley, California 91820, USA.

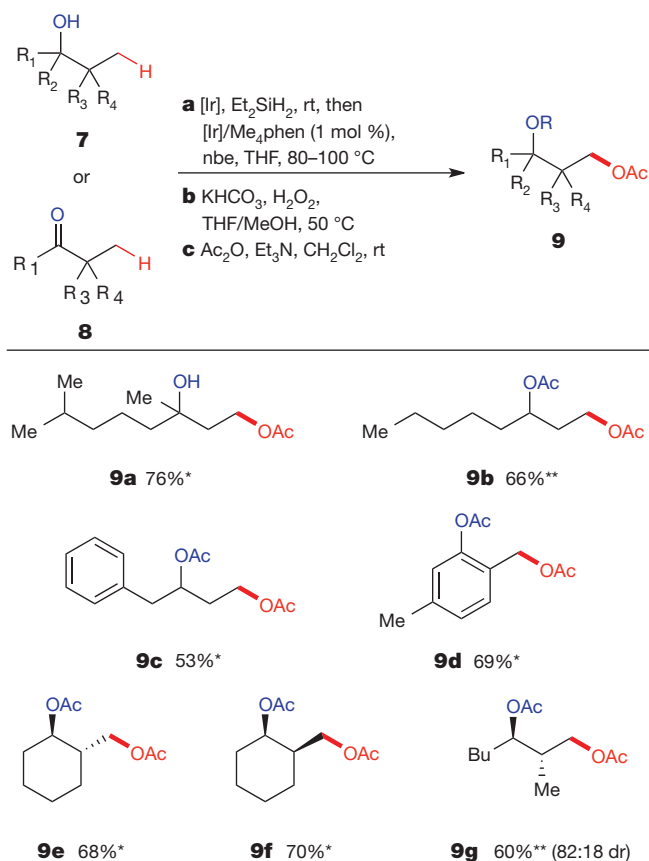


Figure 1 | Hydroxyl-directed γ -oxygenation of secondary and tertiary alcohols and ketones. Overall isolated yields for reactions conducted on a 1.0-mmol scale following purification by silica-gel chromatography. Reagents and conditions: **a**, **7** (to give products marked *) or **8** (products marked **) (1.0 equiv.), Et_2SiH_2 (1.2 equiv.), $[\text{Ir}(\text{cod})\text{OMe}]_2$ (0.05 mol %), THF, room temperature (rt); removal of volatiles, then $[\text{Ir}(\text{cod})\text{OMe}]_2$ (0.5 mol %), Me_4phen (1.2 mol %), *nbe* (1.2 equiv.), THF, 80–100 °C. **b**, KHCO_3 (2.5 equiv.), 30% aqueous H_2O_2 (10 equiv.), THF/MeOH, 50 °C. **c**, Ac_2O (1.5–3.0 equiv.), DMAP (0–0.05 equiv.), $\text{CH}_2\text{Cl}_2/\text{Et}_3\text{N}$, room temperature. Ac_2O , acetic anhydride; DMAP, 4-dimethylaminopyridine; Et_3N , triethylamine. For full experimental details, see the Supplementary Information.

prepared by this sequence of C–H silylation and oxidation. Both tertiary (**9a**) and secondary acyclic alcohols (**9b**, **c**) underwent aliphatic C–H bond oxygenation with comparable efficiency. Under most conditions, oxygenation of a primary C–H bond in the presence of a secondary alcohol would be challenging because of competitive oxidation of the secondary alcohol to the corresponding ketone. Our strategy allows an overall oxidation of a primary C–H bond in the presence of a secondary alcohol because it uses a mildly reducing silane as the functionalizing reagent.

Reactions of phenols (**9d**) and cyclic aliphatic alcohols (**9e**, **f**) also occurred under these conditions in good yields. The C–H functionalization of cyclic alcohols proved insensitive to the stereochemistry of the ring fusion in the bicyclic oxasilolane intermediate; *trans*- and *cis*-2-methylcyclohexanol underwent this process in similar yields (**9e**, **f**). However, the reaction of an acyclic substrate bearing diastereotopic methyl groups formed **9g** as the major product with significant diastereoselectivity (82:18 dr), resulting from C–H bond silylation to form the less strained *trans*-disubstituted oxasilolane intermediate.

The products in Fig. 2 illustrate the range of functional groups that are tolerated by the aliphatic C–H bond oxygenation sequence. Aryl halides (**9h**, **i**), an aryl trifluoromethyl group (**9j**), and an internal carbon–carbon double bond (**9k**) were all compatible with the reaction conditions. The sequence also tolerated phenols protected as the

corresponding benzyl or silyl ethers (**9l–n**); protected catechols (**9o**); and aliphatic alcohols protected with trialkylsilyl groups (**9p**, **q**). Although aldehydes were not compatible, the C–H functionalization occurred in high yield when this functionality was masked as an acetal (**9r**). None of the examples in Fig. 2 suffered from competing intra- or intermolecular aromatic C–H bond functionalization, an observation that is particularly striking for **9l**, which contains three unhindered aromatic C–H bonds. Because the catalyst for the present aliphatic C–H silylation is similar to that used most commonly for aromatic C–H borylation¹⁵, the absence of products from aromatic functionalization can be attributed to the powerful directing ability of the (hydrido)silyl group to promote selective functionalization of a δ -C–H bond.

Whereas a *tert*-butyl ester was tolerated by our standard conditions (**9s**), methyl ester, amide, ketone and carbamate functional groups underwent competitive hydrosilylation when forming the silyl ether in the presence of the iridium catalyst. However, diethyl(hydrido)silyl ethers formed from the coupling of Et_2SiH_2 with alcohols containing auxiliary methyl ester, ketone and carbamate moieties without competitive reduction of the carbonyl functionalities in the presence of 0.2 mol % of the ruthenium complex $\text{Ru}(\text{PPh}_3)_3\text{Cl}_2$ (ref. 17) in which Ph is a phenyl group. Moreover, the carbonyl groups in these compounds were compatible with the C–H silylation process and subsequent Tamao–Fleming oxidation; thus, we obtained diol derivatives **9t**, **9v** and **9w** in good overall yield from the starting alcohols. A small amount of carbonyl hydrosilylation occurred during ruthenium-catalysed silylation of the alcohol containing an amide functional group, but we still obtained diol product **9u** in 52% overall yield.

Because the iridium-catalysed C–H functionalization tolerates auxiliary functionality and is highly selective for reaction at a single C–H bond, it allows site-selective γ -functionalization of natural products bearing hydroxyl and carbonyl groups (Fig. 3). The reaction of (+)-fenchol (**10**) under our standard conditions led to selective functionalization of the C9 methyl group, directed by the C2 hydroxyl group, to generate 9-hydroxyfenchol diacetate (**11a**) in 66% yield (Fig. 3a). Previously, **11a** was accessed indirectly from other terpene skeletons by multi-step rearrangement sequences¹⁸; direct oxidation of (+)-fenchol at the C9 position by fermentation formed diol **11b** in only 4.0% yield and was accompanied by three further diol isomers¹⁹. *Exo*-selective hydrosilylation of (+)-camphor (**12**) catalysed by 0.05 mol % $[\text{Ir}(\text{coe})_2\text{Cl}]_2$, in which *coe* is cyclooctene (92:8 *exo/endo*), followed by directed C–H functionalization at the C10 methyl group, gave **13a** in 57% yield. The previous synthesis of *exo*-2,10-camphanediol (**13b**) began with the significantly more expensive (+)-ketopinacid²⁰.

Finally, we conducted the iridium-catalysed selective functionalization of a pair of triterpenoid saponin aglycons. Triterpenoid saponins (for example, **15**, in which R_1 = sugar, Fig. 3b) exhibit a range of biological activities, including anti-inflammatory, anti-fungal, anti-leishmanial and anti-tumour properties²¹. Together with the carbohydrate portions of saponins, the aglycon moiety is important in modulating biological activity; for example, C23-hydroxylated saponins possess strong haemolytic activity²¹. Because the extraction of saponins from natural sources can be time-consuming and low-yielding, laboratory synthesis from the corresponding aglycon is often the preferred method to gain material for biological studies²¹, and syntheses of both natural and non-natural saponins starting from methyl hederagenate (**15b**) have been described²¹. The potent biological activity of saponins bearing a C23-hydroxylated aglycon, coupled with the broader availability of the C23-unfunctionalized materials, suggest that a general method for chemoselective C23-oxygenation (for example, **14** \rightarrow **15**) would provide a short synthesis of saponins that would otherwise be challenging to access. An earlier synthesis of hederagenin (**15a**) from oleanolic acid (**14a**) required ten steps and proceeded with 36% overall yield²².

As shown in Fig. 3b, our iridium-catalysed silylation of aliphatic C–H bonds converted methyl oleanate (**14b**) to methyl hederagenate

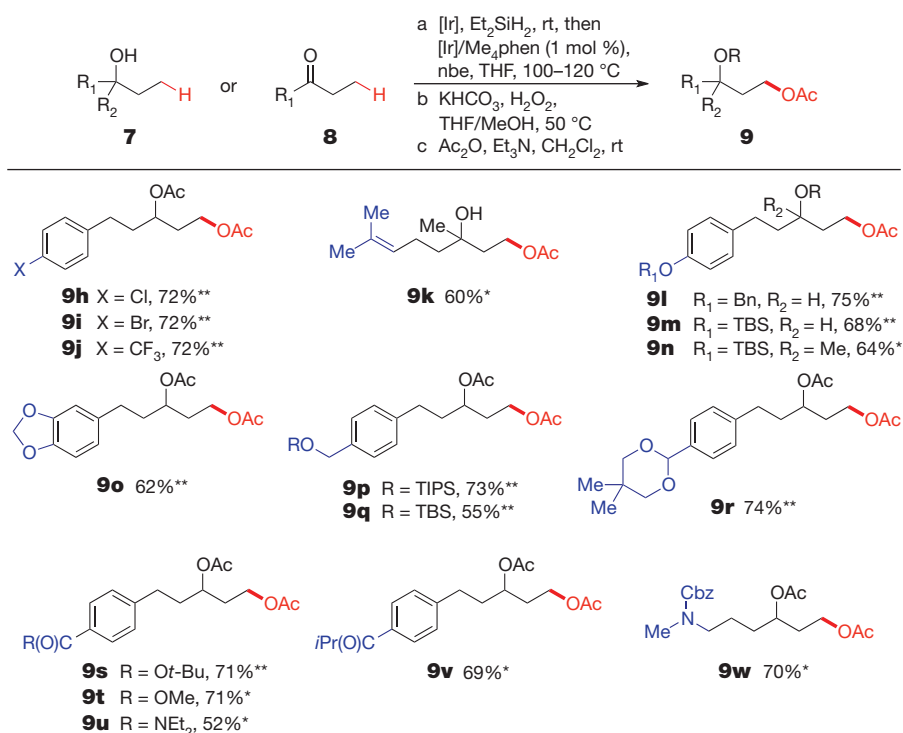


Figure 2 | Functional-group tolerance of hydroxyl-directed γ -oxygenation. Overall isolated yields for reactions conducted on a 1.0-mmol scale following purification by silica-gel chromatography. Reagents and conditions: **a**, **7** (to give products marked ^{*}) or **8** (products marked ^{**}) (1.0 equiv.), Et₂SiH₂ (1.2 equiv.), [Ir(cod)OMe]₂ (0.05 mol %), THF, room temperature; removal of volatiles, then [Ir(cod)OMe]₂ (0.5 mol %), Me₄phen (1.2 mol %), nbe (1.2 equiv.), THF, 100–120 °C. **b**, KHCO₃ (2.5 equiv.), 30% aqueous H₂O₂ (10 equiv.), THF/MeOH, 50 °C. **c**, Ac₂O (1.5–3.0 equiv.), DMAP (0–0.05 equiv.), CH₂Cl₂/Et₃N, room temperature. For compounds **9t**, **9u**, **9v** and **9w**, step **a** (first half) used Ru(PPh₃)₃Cl₂ (0.2 mol %) in benzene or toluene at 50 °C. For compound **9u**, step **a** (second half) used 2 mol % [Ir]/Me₄phen. For full experimental details, see the Supplementary Information.

(**15b**) in three steps and 61% overall yield, without isolation of any intermediates. Similarly, selective oxygenation of the C23 methyl group of methyl glycyrrhetinate (**16**) provided diol **17a** in 51% overall yield. C24-hydroxylated glycyrrhetinate derivative **17b** has previously been obtained from *Glycyrrhiza uralensis*²³, but the C23-hydroxylated

diastereomer **17a** has not been reported. On the basis of these results, we anticipate that this iridium-catalysed C–H bond functionalization should be suitable for the hydroxyl-directed C23-functionalization of a series of triterpenoids, enabling the preparation of both natural and unnatural derivatives.

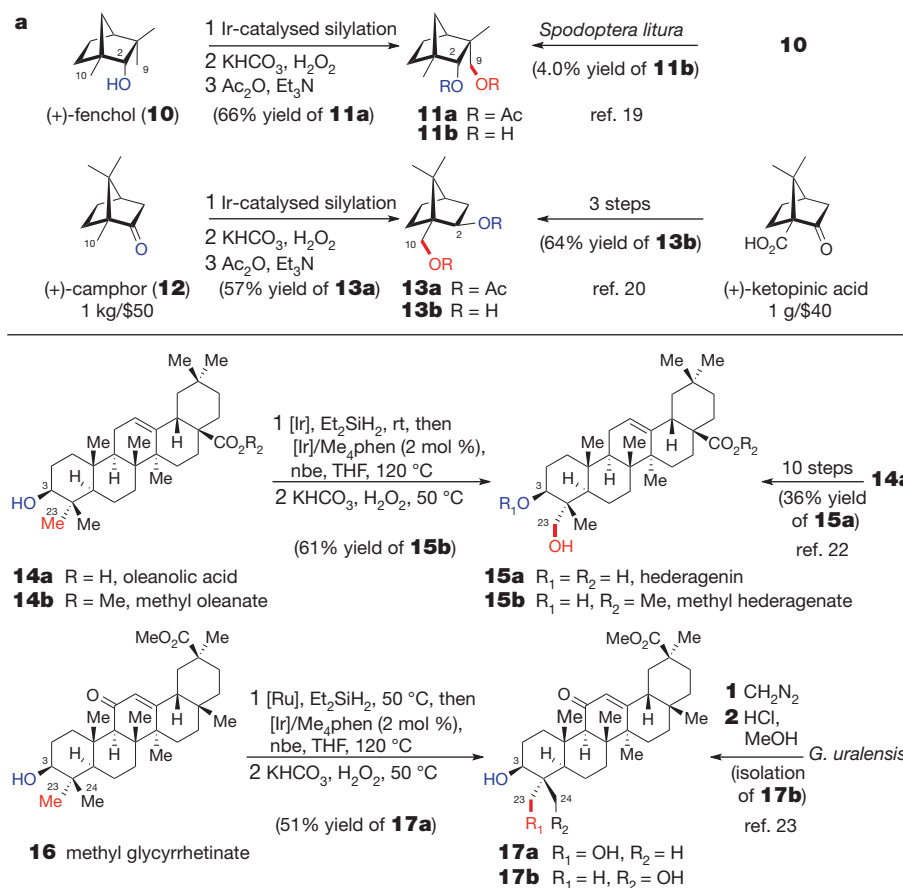


Figure 3 | Directed aliphatic C–H functionalization of natural products.

a, Directed γ -functionalization of the monoterpenes fenchol (**10**) and camphor (**12**). **b**, Hydroxyl-directed γ -functionalization of the triterpenoid saponins methyl oleanate (**14b**) and methyl glycyrrhetinate (**16**). For full experimental details, see the Supplementary Information.

The hydroxyl-directed functionalization of primary C–H bonds that we have developed overcomes several limitations and complements the scope of current methods for aliphatic C–H bond functionalization. The regioselectivity is distinct from that of classical, stoichiometric methods for 1,5-functionalization²⁴, which involve radical intermediates and often require strongly electrophilic reagents; the selectivity for reactions of primary C–H bonds distinguishes our method from more recent alcohol-directed 1,4-functionalizations, which occur only with weaker benzylic and tertiary C–H bonds^{25,26}. Furthermore, by our method, a single reagent docks at the existing functional group and effects the C–H functionalization process. Previous methods for catalytic, directed functionalization of aliphatic C–H bonds, which inspired the design of the current system, have led to regioselective functionalization of neighbouring C–H bonds but require installation and removal of a separate directing group before and after the C–H bond functionalization and purification of the intermediates^{27,28}. Finally, the iridium catalyst reacts with primary C–H bonds in preference to secondary C–H bonds, and this selectivity complements that of non-directed oxidations that typically occur at the weaker and more electron-rich C–H bonds^{9,10,29}. Considering the high selectivity of this process, the tolerance for auxiliary functionality, and the complementarity with radical-based reactions and direct oxidations, this method is suitable for both *de novo* synthesis of complex organic molecules and derivatization of natural products. It should inspire the development of related strategies for selective C–H bond functionalization with reagents that both dock at existing functional groups and deliver functionality to a single C–H bond³⁰.

METHODS SUMMARY

We assembled iridium-catalysed silylation reactions under a nitrogen atmosphere using oven-dried glassware and dry, deoxygenated solvents. We assembled Tamao–Fleming oxidation reactions under air using commercial-grade solvents. We monitored reactions using gas chromatography (with a mass spectral detector or a flame ionization detector) or ¹H nuclear magnetic resonance (NMR) spectroscopy.

General procedure for iridium-catalysed, directed aliphatic C–H bond functionalization. We dissolved an alcohol or ketone substrate in tetrahydrofuran (THF) and treated it with a freshly prepared solution of [Ir(cod)OMe]₂ (0.05 mol %) in THF and then with neat Et₂SiH₂ (1.2 equiv.). We stirred the resulting solution at room temperature (23 °C) until we observed complete conversion of the alcohol or ketone to the corresponding diethyl(hydrido)silyl ether. We then placed the reaction mixture under high vacuum for 1 h. We treated the concentrated diethyl(hydrido)silyl ether sequentially with freshly prepared solutions of nbe (1.2 equiv.) in THF and [Ir(cod)OMe]₂ (0.5 mol %) in THF, and then with a slurry of Me₃phen (1.25 mol %) in THF. We stirred the resulting solution at room temperature for 1 h and then heated it at 80–120 °C until we observed complete conversion to the corresponding oxasilolane. We then treated the crude reaction mixture containing the oxasilolane sequentially with MeOH, KHCO₃ (2.5 equiv.) and H₂O₂ (30% solution in H₂O, 10 equiv.), and stirred the resulting mixture overnight at 50 °C. We carefully quenched the reaction with aqueous NaHSO₃, and extracted the resulting mixture with EtOAc. We washed the combined organic layers sequentially with 1 M HCl and saturated NaHCO₃, and then dried them with MgSO₄. We filtered the resulting organic layer through Celite and concentrated it to provide the crude diol, which we either purified directly or converted to the corresponding acetate derivative through treatment with Ac₂O and Et₃N. For full experimental details and characterization of all new compounds, along with copies of ¹H and ¹³C NMR spectra, see Supplementary Information.

Received 29 August; accepted 13 December 2011.

1. Gutekunst, W. R. & Baran, P. S. C–H functionalization logic in total synthesis. *Chem. Soc. Rev.* **40**, 1976–1991 (2011).
2. McMurray, L., O'Hara, F. & Gaunt, M. J. Recent developments in natural product synthesis using metal-catalysed C–H bond functionalisation. *Chem. Soc. Rev.* **40**, 1885–1898 (2011).
3. Chen, K. & Baran, P. S. Total synthesis of eudesmane terpenes by site-selective C–H oxidations. *Nature* **459**, 824–828 (2009).
4. Godula, K. & Sames, D. C–H bond functionalization in complex organic synthesis. *Science* **312**, 67–72 (2006).

5. Feng, Y. & Chen, G. Total synthesis of celogentin C by stereoselective C–H activation. *Angew. Chem. Int. Ed.* **49**, 958–961 (2010).
6. Giannis, A., Heretsch, P., Sarli, V. & Stöbel, A. Synthesis of cyclopamine using a biomimetic and diastereoselective approach. *Angew. Chem. Int. Ed.* **48**, 7911–7914 (2009).
7. Yu, J.-Q. & Shi, Z. in *Topics in Current Chemistry* Vol. 292 (Springer, 2010).
8. Lyons, T. W. & Sanford, M. S. Palladium-catalyzed ligand-directed C–H functionalization reactions. *Chem. Rev.* **110**, 1147–1169 (2010).
9. Chen, M. S. & White, M. C. A predictably selective aliphatic C–H oxidation reaction for complex molecule synthesis. *Science* **318**, 783–787 (2007).
10. Newhouse, T. & Baran, P. S. If C–H bonds could talk: selective C–H bond oxidation. *Angew. Chem. Int. Ed.* **50**, 3362–3374 (2011).
11. Lu, Y., Wang, D.-H., Engle, K. M. & Yu, J.-Q. Pd(II)-catalyzed hydroxyl-directed C–H olefination enabled by monoprotected amino acid ligands. *J. Am. Chem. Soc.* **132**, 5916–5921 (2010).
12. Henkel, T., Brunne, R. M., Müller, H. & Reichel, F. Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem. Int. Ed.* **38**, 643–647 (1999).
13. Simmons, E. M. & Hartwig, J. F. Iridium-catalyzed arene *ortho*-silylation by formal hydroxyl-directed C–H activation. *J. Am. Chem. Soc.* **132**, 17092–17095 (2010).
14. Boebel, T. A. & Hartwig, J. F. Silyl-directed, iridium-catalyzed *ortho*-borylation of arenes. A one-pot *ortho*-borylation of phenols, arylamines, and alkylarenes. *J. Am. Chem. Soc.* **130**, 7534–7535 (2008).
15. Mkhali, I. A. I., Barnard, J. H., Marder, T. B., Murphy, J. M. & Hartwig, J. F. C–H activation for the construction of C–B bonds. *Chem. Rev.* **110**, 890–931 (2010).
16. Crotti, C. *et al.* Evaluation of the donor ability of phenanthrolines in iridium complexes by means of synchrotron radiation photoemission spectroscopy and DFT calculations. *Dalton Trans.* 133–142 (2007).
17. Corriu, R. J. P. & Moreau, J. J. E. Selective catalytic route to bifunctional silanes. Catalysis by rhodium and ruthenium complexes of the alcoholysis of diarylsilanes and the hydrosilylation of carbonyl compounds. *J. Chem. Soc. Chem. Commun.* 38–39 (1973).
18. Bosworth, N. & Magnus, P. D. Studies on terpenes. Part I. Rearrangement of 7-oxatricyclo[4.3.0.0]nonanes into 8-substituted 1,3,3-trimethylnorbornane derivatives. *J. Chem. Soc. Perkin Trans. 1* 943–948 (1972).
19. Miyazawa, M. & Miyamoto, Y. Biotransformation of (+)-(1R,2S)-fenchol by the larvae of common cutworm (*Spodoptera litura*). *Tetrahedron* **60**, 3091–3096 (2004).
20. Deng, J. G., Jiang, Y. Z., Liu, G. L., Wu, L. J. & Mi, A. Q. A practical method for the synthesis of homochiral 2,10-camphanediols. *Synthesis* 963–965 (1991).
21. Plé, K., Chwalek, M. & Voutquenne-Nazabadioko, L. Synthesis of α -hederin, δ -hederin, and related triterpenoid saponins. *Eur. J. Org. Chem.* **2004**, 1588–1603 (2004).
22. García-Granados, A., López, P. E., Melguizo, E., Parra, A. & Simeó, Y. Remote hydroxylation of methyl groups by regioselective cyclopalladation. Partial synthesis of hyptatic acid-A. *J. Org. Chem.* **72**, 3500–3509 (2007).
23. Kitagawa, I., Hori, K., Sakagami, M., Zhou, J. L. & Yoshikawa, M. Saponin and sapogenol. XLVIII. On the constituents of the roots of *Glycyrrhiza uralensis* Fischer from northeastern China. (2). Licorice-saponins D3, E2, F3, G2, H2, J2, and K2. *Chem. Pharm. Bull. (Tokyo)* **41**, 1337–1345 (1993).
24. Majetich, G. & Wheelless, K. Remote intramolecular free radical functionalizations: an update. *Tetrahedron* **51**, 7095–7129 (1995).
25. Chen, K., Richter, J. M. & Baran, P. S. 1,3-diol synthesis via controlled, radical-mediated C–H functionalization. *J. Am. Chem. Soc.* **130**, 7247–7249 (2008).
26. Kasuya, S., Kamijo, S. & Inoue, M. Direct construction of 1,3-diaxial diol derivatives by C–H hydroxylation. *Org. Lett.* **11**, 3630–3632 (2009).
27. Desai, L. V., Hull, K. L. & Sanford, M. S. Palladium-catalyzed oxygenation of unactivated sp³ C–H bonds. *J. Am. Chem. Soc.* **126**, 9542–9543 (2004).
28. Giri, R. *et al.* Pd-catalyzed stereoselective oxidation of methyl groups by inexpensive oxidants under mild conditions: a dual role for carboxylic anhydrides in catalytic C–H bond oxidation. *Angew. Chem. Int. Ed.* **44**, 7420–7424 (2005).
29. Litvinas, N. D., Brodsky, B. H. & Du Bois, J. C–H hydroxylation using a heterocyclic catalyst and aqueous H₂O₂. *Angew. Chem. Int. Ed.* **48**, 4513–4516 (2009).
30. Zalatan, D. N. & Du Bois, J. in *Topics in Current Chemistry* Vol. 292 (eds Yu, J.-Q. & Shi, Z.) 347–378 (Springer, 2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the US National Science Foundation (CHE-0910641 to J.F.H.) and the US National Institutes of Health (GM087901 to E.M.S.) for funding this work, and Johnson Matthey for a gift of [Ir(cod)OMe]₂.

Author Contributions E.M.S. and J.F.H. conceived the work and designed the experiments. E.M.S. performed the experiments. Both authors analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.F.H. (jhartwig@berkeley.edu).

Opposite effects of fear conditioning and extinction on dendritic spine remodelling

Cora Sau Wan Lai¹, Thomas F. Franke² & Wen-Biao Gan¹

It is generally believed that fear extinction is a form of new learning that inhibits rather than erases previously acquired fear memories^{1–3}. Although this view has gained much support from behavioural and electrophysiological studies^{1–10}, the hypothesis that extinction causes the partial erasure of fear memories remains viable. Using transcranial two-photon microscopy^{11,12}, we investigated how neural circuits are modified by fear learning and extinction by examining the formation and elimination of postsynaptic dendritic spines of layer-V pyramidal neurons in the mouse frontal association cortex. Here we show that fear conditioning by pairing an auditory cue with a footshock increases the rate of spine elimination. By contrast, fear extinction by repeated presentation of the same auditory cue without a footshock increases the rate of spine formation. The degrees of spine remodelling induced by fear conditioning and extinction strongly correlate with the expression and extinction of conditioned fear responses, respectively. Notably, spine elimination and formation induced by fear conditioning and extinction occur on the same dendritic branches in a cue- and location-specific manner: cue-specific extinction causes formation of dendritic spines within a distance of two micrometres from spines that were eliminated after fear conditioning. Furthermore, reconditioning preferentially induces elimination of dendritic spines that were formed after extinction. Thus, within vastly complex neuronal networks, fear conditioning, extinction and reconditioning lead to opposing changes at the level of individual synapses. These findings also suggest that fear memory traces are partially erased after extinction.

Classical fear conditioning is widely used to study associative learning in which a conditioned neutral stimulus (CS; for example an auditory cue) is paired with the presentation of an unconditioned aversive stimulus (US; for example a footshock) to elicit a conditioned response¹³ (CR; for example a freezing response to CS in the absence of US). Repeated exposures to CS diminish the expression of the CR, a process called extinction^{1–3}. It is widely believed that fear extinction involves new learning of the 'safe' association between CS and the absence of US, rather than an erasure of the original CS–US association^{1–3}. This theory is supported by behavioural studies of spontaneous recovery, renewal and reinstatement of fear memory after extinction^{1–6}. Furthermore, fear conditioning and extinction regulate activities of non-overlapping neuronal populations in the amygdala, hippocampus and frontolimbic cortex^{7–10}. Although these behavioural and electrophysiological studies suggest that fear memories are stored in different circuits from extinction memories, they do not exclude the possibility that extinction may cause a partial erasure of fear memory traces. In support of the second view, it has been found that spontaneous recovery of fear memories is minimal after extinction training in young animals¹⁴. Furthermore, fear conditioning and extinction are accompanied by an increase or decrease in synaptic activity in amygdala¹⁵ and in the expression level of signalling molecules involved in memory formation^{16–18}.

To investigate how fear conditioning and extinction affect synaptic circuits to result in opposite behavioural responses, we used transcranial

two-photon microscopy to examine the formation and elimination of postsynaptic dendritic spines of layer-V pyramidal neurons in the dorsal medial region of the frontal association cortex^{11,12,19} (FrA; Fig. 1a). We chose FrA to investigate synaptic changes associated with fear conditioning and extinction for the following reasons. First, consistent with previous reports in rats^{20,21}, we observed reciprocal connections between the mouse FrA and amygdala (Fig. 1b, c and Supplementary Figs 1 and 2), suggesting that this cortical region interacts directly with the amygdala to participate in fear learning and extinction. Second, consistent with findings in rats that FrA is important for fear memory consolidation²², we found that after fear conditioning or extinction, inactivation of the mouse FrA by muscimol injection impaired the consolidation of fear and extinction memories (Supplementary Fig. 3). Lastly, unlike prelimbic and infralimbic regions that are important for fear expression and extinction but located deep in the brain^{7,8,23}, FrA is accessible for *in vivo* two-photon imaging of dendritic spine plasticity.

Using two-photon microscopy and YFP-expressing transgenic mice^{11,12}, we first examined whether fear conditioning affected spine formation and elimination over 48 h in FrA (Fig. 1d, f). In these experiments, one-month-old mice were imaged and then subjected to one of four stimulus conditions: three tones each paired with a co-terminating footshock; three tones each temporally unpaired from a footshock; three tones only; or three footshocks only. Forty-eight hours after fear conditioning, we put mice through a tone-cued recall test to assess conditioned freezing responses, followed by a second imaging session to examine spine dynamics. We found that only the CS–US paired group, but none of the other groups, showed robust freezing responses during the recall test ($F_{(3,16)} = 18.689$, $P < 0.001$; Fig. 1e). Notably, only the paired group showed a significant increase in spine elimination after 48 h when compared with the unpaired, tone-only or shock-only groups ($17.4 \pm 2.3\%$ versus $9.7 \pm 1.9\%$, $10.1 \pm 1.1\%$ or $10.0\% \pm 1.3\%$, respectively; $F_{(3,16)} = 24.569$, $P < 0.001$; Fig. 1f; see also Supplementary Fig. 4). We did not observe a significant difference in spine formation among the four groups ($F_{(3,16)} = 0.164$, $P > 0.9$; Fig. 1f and Supplementary Fig. 4). There were also no significant differences in the formation and elimination rates of dendritic filopodia, that is, precursors of dendritic spines²⁴, among different groups (Supplementary Fig. 5). These results indicate that fear conditioning by CS–US association causes dendritic spine elimination over 2 d in FrA.

To investigate the impact of fear conditioning further, we examined freezing responses and spine dynamics in the paired and unpaired groups 9 d after conditioning. A recall test showed that mice subjected to the paired CS–US stimuli showed a higher level of conditioned freezing responses than mice exposed to the unpaired CS–US stimuli ($P < 0.001$; Fig. 1g). Furthermore, spine elimination over 9 d was significantly higher in the paired group than in the unpaired control group ($P < 0.05$; Fig. 1h). No significant differences in spine formation were observed between the two groups ($P > 0.05$; Fig. 1h). Thus, spine elimination induced by fear conditioning represents a long-lasting synaptic change in FrA. In addition, we found a significant increase in spine elimination but not formation as early as 24 h after fear

¹Molecular Neurobiology Program, Skirball Institute, Department of Physiology and Neuroscience, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA.

²Departments of Psychiatry and Pharmacology, New York University School of Medicine, 550 First Avenue, New York, New York 10016, USA.

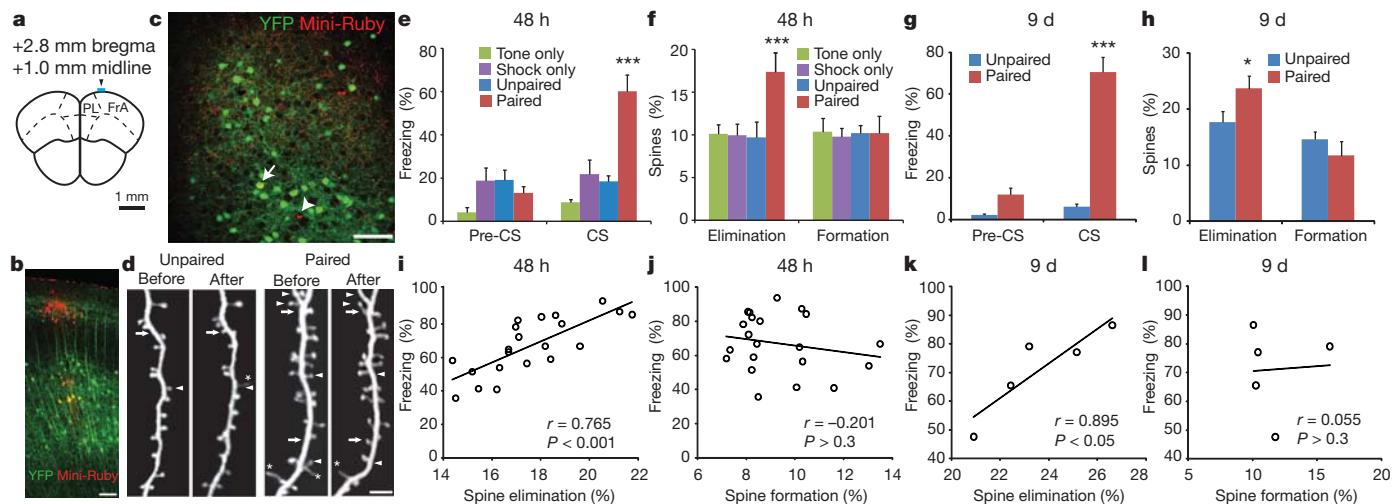


Figure 1 | Fear conditioning causes spine elimination. **a**, Diagram of a coronal section of frontal association cortex (FrA) showing the imaging site (cyan bar). PL, prelimbic cortex. **b**, **c**, Neurons labelled with Mini-Ruby dye (**b**, red) in FrA have extensive axonal arborizations in amygdala (**c**, red). The arrow and arrowhead in **c** respectively indicate yellow fluorescent protein (YFP)-positive and -negative neuronal soma labelled with Mini-Ruby in amygdala (see also Supplementary Fig. 1). **d**, Representative images of dendrites before and after conditioning in the unpaired and paired groups. Arrows and arrowheads indicate spine formation and elimination, respectively. Asterisks mark filopodia. **e**, Percentage of freezing in different groups before and during

CS presentation 48 h after conditioning. **f**, Percentage of spine elimination and formation 48 h after conditioning ($n = 5$ for each group). Only the paired group showed an increase in freezing response (**e**) and spine elimination (**f**). **g**, Freezing responses of the paired and unpaired groups 9 d after conditioning. **h**, Percentage of spine elimination and formation over 9 d. (Unpaired, $n = 4$; paired, $n = 5$). **i–l** Freezing response correlated with spine elimination but not formation over either 48 h (**i**, **j**) or 9 d (**k**, **l**). Each circle in **i–l** represents an animal. *** $P < 0.001$, * $P < 0.05$. Data show mean \pm s.e.m. (**e**, **g**) or mean \pm s.d. (**f**, **h**). Scale bars: 100 μ m (**b**, **c**); 4 μ m (**d**).

conditioning in FrA (Supplementary Fig. 6a). However, no significant increases in dendritic spine elimination or formation were observed 24 h after fear conditioning in the barrel cortex (Supplementary Fig. 6b). Together, these results indicate that fear conditioning causes rapid and long-lasting spine elimination in FrA, but not in all cortical regions.

Notably, we observed that, 48 h after fear conditioning, the percentage of spine elimination correlated significantly with the degree of freezing responses to CS ($r = 0.765$ (correlation coefficient), $P < 0.001$; Fig. 1i). By contrast, there was no significant correlation between the degree of spine formation and freezing responses ($r = -0.201$, $P > 0.3$; Fig. 1j). Furthermore, 9 d after training, freezing responses in the paired group correlated significantly with spine elimination ($r = 0.895$, $P < 0.05$; Fig. 1k) but not spine formation ($r = 0.055$, $P > 0.3$; Fig. 1l). These findings suggest that spine elimination induced by fear conditioning represents an important synaptic change that strongly predicts the conditioned freezing response.

To determine whether fear extinction affects spine dynamics, we subjected mice to fear conditioning as before, followed by 2 d of extinction training through repeated presentation of CS in the absence of footshocks (five trials per day) (Fig. 2a). Dendritic spines in FrA were imaged before and after extinction training, and spine dynamics were compared between fear-conditioned mice with and without extinction. Consistent with the reported effectiveness of our extinction protocol⁷, the conditioned freezing response was significantly reduced in the extinction group ($P < 0.001$) but not in the no-extinction group ($P > 0.3$; Fig. 2b). After 2 d of extinction, we found that spine formation was significantly higher in the extinction group than in the no-extinction group ($16.5 \pm 2.6\%$ versus $8.3 \pm 1.2\%$, $P < 0.001$; Fig. 2c). By contrast, no significant difference in spine elimination was observed after extinction ($P > 0.2$; Fig. 2c). There were also no significant differences in the formation and elimination rates of dendritic filopodia after extinction (Supplementary Fig. 7). Notably, the freezing response to CS after extinction showed a significant inverse correlation with spine formation ($r = -0.809$, $P < 0.01$; Fig. 2d) but not with spine elimination ($r = -0.250$, $P > 0.4$; Fig. 2e). Thus, by contrast with fear conditioning, extinction induces new spine formation, the degree of which predicts the effectiveness of extinction training in reducing conditioned freezing responses.

A recent study has shown that a fraction of new spines induced by novel sensory and motor learning experiences persist over months¹². To determine whether new spines induced by extinction are long lasting, we subjected mice to fear conditioning followed by extinction training as described above. On day 12, 7 d after extinction training, we performed a recall test to assess the extinction memory and then re-imaged mice to determine the persistence of spines formed over the 2-d extinction training (Fig. 2a). We found that the acquired extinction memory was intact 7 d after extinction training (Fig. 2f). Furthermore, the percentage of new spines persisting until day 12 was inversely correlated with the freezing response to the CS ($r = -0.899$, $P < 0.05$; Fig. 2g). Thus, long-lasting new spines induced by extinction training may contribute to the preservation of the extinction memory.

Our results so far indicate that fear conditioning predominantly promotes spine elimination whereas extinction mainly induces spine formation. To determine whether fear conditioning and extinction cause spine remodelling on the same or different dendritic branches, we measured the percentage of spine elimination after fear conditioning and the percentage of spine formation after extinction along individual dendritic branches 15–50 μ m in length (average length, $27.6 \pm 9.3 \mu$ m). We found that the percentage of spine elimination after fear conditioning positively correlated with the percentage of spine formation after extinction ('CS/tone-A extinction'; $n = 6$ mice, 40 branches, $r = 0.458$, $P < 0.01$; Fig. 3b, e). By contrast, no significant correlation was observed in the no-extinction group ($n = 5$ mice, 25 branches, $r = 0.240$, $P > 0.2$; Fig. 3a, d). Furthermore, when a different cue (tone B) was presented repeatedly instead of the cue originally used for fear conditioning (tone A), no significant correlation was observed between spine elimination and formation ('tone B'; $n = 5$ mice, 41 branches, $r = 0.231$, $P > 0.1$; Fig. 3c, f). These results suggest that cue-specific extinction induces spine formation on the dendritic branches on which fear conditioning previously caused spine elimination.

Further to understand how fear conditioning and extinction impact synaptic connectivity on the same dendritic branches, we measured the distance between the site of conditioning-induced spine elimination and the closest site of extinction-induced spine formation. In the extinction group, $57.3 \pm 5.5\%$ of extinction-induced new spines were located within 2 μ m of either side of a spine that had previously been

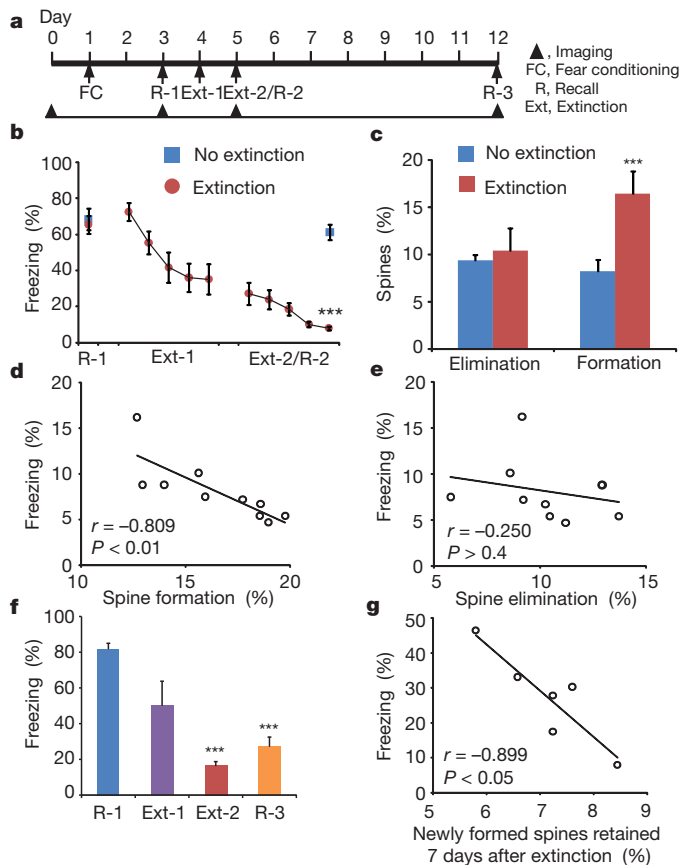


Figure 2 | Fear extinction induces spine formation. **a**, Timeline of experimental manipulations and imaging. **b**, Percentage of freezing during the recall test and extinction training. Freezing in the extinction group ($n = 10$) was lower than the no-extinction group ($n = 5$, $P < 0.001$). **c**, Extinction significantly increased spine formation ($P < 0.001$). **d, e**, Percentage of freezing after extinction was inversely correlated with spine formation (d) but not elimination (e). Extinction training was performed on two consecutive days (Ext-1 and Ext-2). Freezing in the last trial of Ext-2 in b was plotted in d and e. **f**, Percentage of freezing in the recall tests and during the last trials of extinction on days 4 and 5. Decreased freezing after extinction persisted for 7 d as shown in the recall test (R-3). There was no significant difference in freezing between the last trial of Ext-2 and R-3 ($n = 6$, $P > 0.1$). **g**, New spines induced by extinction and persisting on day 7 inversely correlated with freezing response. $***P < 0.001$. Data show mean \pm s.e.m. (b, f) or mean \pm s.d. (c).

eliminated after conditioning. In the groups that did not undergo extinction or were exposed to tone B instead of the conditioned stimulus (CS/tone A), only $24.9 \pm 1.4\%$ or $24.4 \pm 4.9\%$ of new spines, respectively, were formed within $2 \mu\text{m}$ of spines previously eliminated by conditioning (Fig. 3g). Moreover, within a distance of $2 \mu\text{m}$ from sites of spine elimination, $\sim 80\%$ of newly formed spines in all groups were oriented within 90° of previously eliminated spines. Beyond $2 \mu\text{m}$, the likelihood of newly formed spines being oriented within 90° of previously eliminated spines decreased to chance levels ($\sim 50\%$) (Fig. 3h). These results indicate that after extinction training, new spines tend to form within close proximity and orient in the same direction as spines that were eliminated after fear conditioning. Furthermore, we estimated that the number of presynaptic boutons available for a synaptic contact with a spine that is $2 \mu\text{m}$ long and located along a $4\text{-}\mu\text{m}$ dendritic segment is ~ 32 (Supplementary Information, section 1). When considering that new spines tend to be oriented in the same direction as previously eliminated spines, our results indicate that there could be a $\sim 1/16$ chance that new spines will contact the same synaptic boutons as previously eliminated spines.

So far, our data have indicated that extinction with the US-associated conditioned stimulus (tone A), but not with an unconditioned stimulus

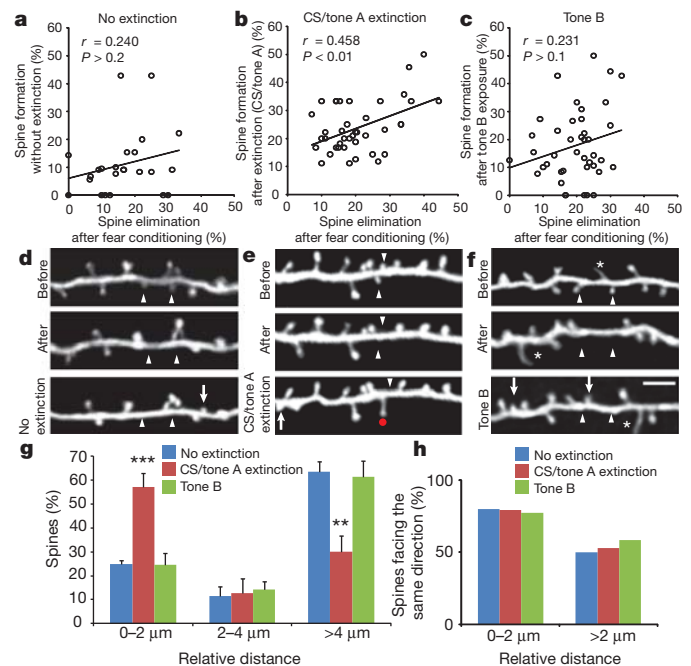


Figure 3 | Fear conditioning and extinction cause location-specific spine remodelling. **a-c**, Correlations between spine elimination 48 h after fear conditioning and spine formation on individual dendritic branches under the following three conditions: no extinction training (a); after extinction training using the conditioned tone (b); after repeated exposure to a novel tone (c). **d-f**, Representative images of dendritic branches taken before and after conditioning from the three groups. Arrowheads mark sites of spine elimination induced by conditioning. Arrows mark sites of newly formed spines under the three conditions. The red dot in e marks a new spine formed after extinction and located within $2 \mu\text{m}$ of a spine previously eliminated after conditioning. Asterisks mark filopodia. Scale bar in f, $4 \mu\text{m}$. **g**, Spine distribution graph depicting the relative distance between sites of spine elimination induced by conditioning and the respective closest sites of spine formation under the three conditions. **h**, Percentage of newly formed spines facing the same direction as previously eliminated spines (oriented within 90° relative to the eliminated spines). $***P < 0.001$, $**P < 0.01$. Data show mean \pm s.e.m. (g).

(tone B), induces spine formation in close proximity to spines that were eliminated after fear conditioning. Further to investigate the cue specificity and location specificity of extinction-induced spine formation, we fear-conditioned mice to tone A (CS1) and tone B (CS2) in two consecutive training sessions using three pairings of CS1 with US and CS2 with US (Fig. 4a). Eliminated spines induced by CS1-US or CS2-US pairings were identified separately over 4 d. In recall tests on day 2 (R-1) and day 4 (R-2), mice showed freezing responses to both CS1 and CS2. We then subjected mice to extinction training with CS2 for another 2 d and newly formed spines induced by extinction were identified (Fig. 4a). On day 7, after CS2 extinction, mice showed a low freezing response to CS2 but a high freezing response to CS1, demonstrating the cue specificity of our extinction training (Fig. 4b). Consistent with our results in Figs 1 and 2, fear conditioning with both CS1 and CS2 increased spine elimination whereas extinction with CS2 promoted spine formation (Fig. 4c). Notably, a significantly larger population of newly formed spines induced by extinction with CS2 were located within $2 \mu\text{m}$ of spines that were eliminated by CS2-US ('CS2 elim/CS2 form') than by CS1-US ('CS1 elim/CS2 form') ($45.9 \pm 7.3\%$ versus $26.6 \pm 3.8\%$, $P < 0.05$; Fig. 4d). These findings indicate that extinction training with a specific auditory cue induces spine formation in close proximity to spines previously eliminated after fear conditioning against the same cue.

Further to investigate the cue and location specificity of opposing synaptic changes after fear conditioning and extinction, we tested

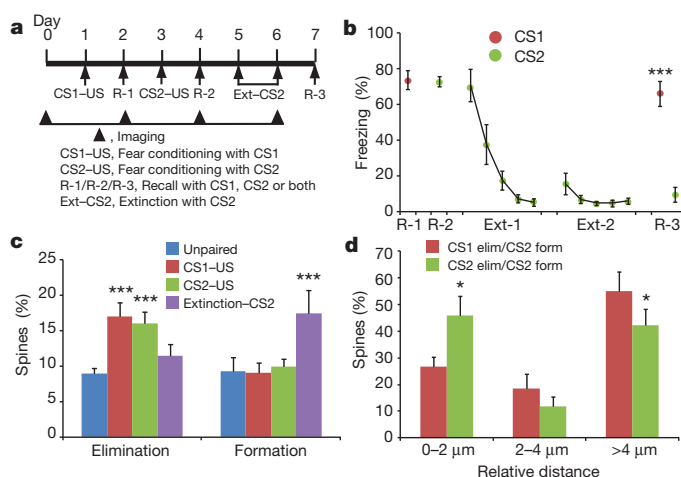


Figure 4 | Extinction induces spine formation in a cue- and location-specific manner. **a**, Timeline of experimental manipulations and imaging. **b**, Percentage of freezing during the recall test and extinction. After extinction with CS2, mice showed a lower freezing response to CS2 than to CS1 on day 7 ($P < 0.001$). **c**, Conditioning to either CS1 or CS2 increased spine elimination when compared with unpaired controls ($P < 0.001$). Extinction with CS2 increased spine formation when compared with the unpaired controls, CS1-US or CS2-US ($F_{(3,16)} = 18.540$, $P < 0.001$). **d**, Spine distribution graph depicting the relative distance between new spines induced by CS2 extinction and eliminated spines induced by CS1-US or CS2-US. A significantly larger population of new spines induced by CS2 extinction were located within 2 μm of spines previously eliminated by CS2-US than were located within 2 μm of spines previously eliminated by CS1-US ($n = 5$ mice, 36 branches, 74 spines, $P < 0.05$). *** $P < 0.001$, * $P < 0.05$. Data show mean \pm s.e.m. (**b**, **d**) or mean \pm s.d. (**c**).

whether new spines induced by extinction are selectively eliminated in a cue-specific manner after reconditioning. In this experiment, we first subjected mice to fear conditioning and extinction, and then reconditioned them by re-exposure to five CS1-US (CS1/tone A) pairings on days 5 and 6 (Fig. 5a). As control groups, mice were either subjected to five temporally unpaired presentations of CS1 and US stimuli or conditioned with five pairings of CS2-US pairings (CS2/tone B) on days 5 and 6 (Fig. 5a–d). Recall tests showed a significant increase of the freezing response to CS1 in the reconditioning group when compared with the unpaired group or the group conditioned with CS2-US ($P < 0.01$; Fig. 5e). We next compared the respective persistences of newly formed spines located within 2 μm or more than 2 μm from previously eliminated spines in the three groups. The persistence of extinction-induced new spines within the 2-μm boundaries in the reconditioning group ($11.1 \pm 1.8\%$) was significantly lower than that in the unpaired or CS2-US conditioned groups ($32.0 \pm 6.2\%$ and $35.2 \pm 11.3\%$, respectively) ($P < 0.05$). There was no significant difference in the persistence of spines formed outside the 2-μm boundary among the three groups ($P > 0.6$; Fig. 5f). Thus, fear reconditioning opposes the effects of extinction on synaptic remodelling with high anatomical and cue specificity.

It is well accepted that synaptic reorganization is critical for learning and memory^{12,13}. However, it is unclear how synaptic circuits are modified by opposing forms of learning and how such modifications contribute to opposite behavioural outcomes. Our studies indicate that fear conditioning, extinction and reconditioning cause opposing synaptic modifications on the same dendritic branches in a cue- and location-specific manner. These findings also suggest that extinction induces at least partial erasure of fear memory traces in FrA. FrA in rodents has reciprocal connections with multiple brain areas including the amygdala^{20,21}, and its inactivation impairs fear learning and extinction²² (Fig. 1b, c and Supplementary Figs 1–3), suggesting that this region is directly involved in modulating freezing behaviours. Future investigations are needed to understand the mechanisms underlying

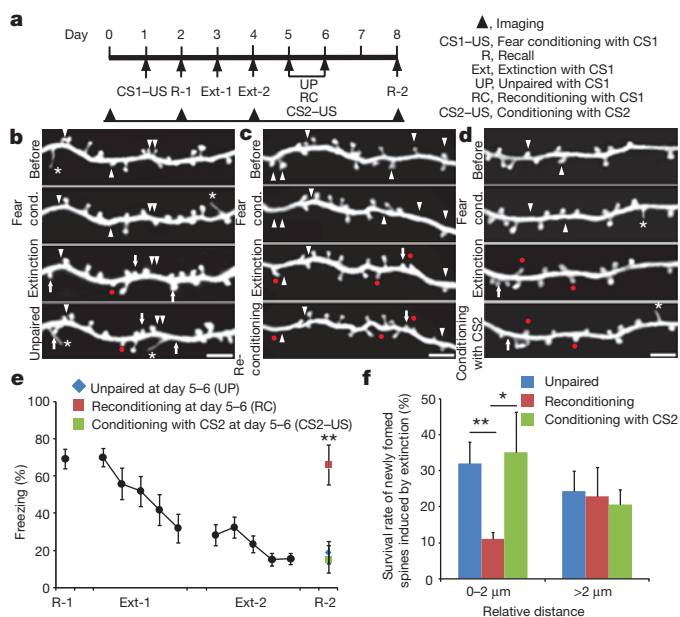


Figure 5 | Reconditioning eliminates spines formed during extinction.

a, Timeline of experimental manipulations and imaging. **b–d**, Representative images before and after fear conditioning, after extinction and after presentation of the unpaired stimuli (**b**), after reconditioning to CS1 (**c**) or after new conditioning to CS2 (**d**). Arrowheads mark sites of spine elimination after fear conditioning. Arrows and red dots mark new spines that were formed after extinction and located more than 2 μm (arrows) and within 2 μm distance (red dots) from previously eliminated spines after conditioning. Asterisks mark filopodia. Scale bar, 4 μm. **e**, Reconditioning (RC) increased freezing by comparison with unpaired stimuli (UP) or new conditioning to CS2 (CS2-US) during recall test R-2 with CS1 (** $P < 0.01$). **f**, The survival rate of new spines induced by extinction training and located within 2 μm of previously eliminated spines was significantly lower after RC-CS1 (54 branches, 107 total new spines) than after UP (58 branches, 107 total new spines) or CS2-US (54 branches, 94 total new spines) (** $P < 0.01$, * $P < 0.05$). Data show mean \pm s.e.m. (**e**) or mean \pm s.d. (**f**).

the opposite changes of synaptic connections in FrA and how such changes contribute to the acquisition, extinction and reinstatement of fear memories.

METHODS SUMMARY

One-month-old male mice expressing YFP (H-line) were used in this study. Fear conditioning was conducted with three pairings of a 30-s, 80-dB auditory cue (400 Hz; tone A = CS1 or 1,200 Hz, tone B = CS2) co-terminating with a 2-s, 0.5-mA scrambled footshock (US). Extinction training was conducted with five CS presentations (each CS lasting 2 min with an intertrial interval of 2 min) per day for 2 d. The procedures of *in vivo* transcranial two-photon imaging and data quantification were as described previously^{11,12}. Either analysis of variance or Student's *t*-test was used to compare spine remodelling and freezing responses among different groups. The Pearson correlation coefficient was used to measure the strength of linear dependence between different variables.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 July; accepted 9 December 2011.

Published online 19 February 2012.

- Myers, K. M. & Davis, M. Mechanisms of fear extinction. *Mol. Psychiatry* **12**, 120–150 (2007).
- Quirk, G. J. & Mueller, D. Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* **33**, 56–72 (2008).
- Bouton, M. E. Context and behavioral processes in extinction. *Learn. Mem.* **11**, 485–494 (2004).
- Rescorla, R. A. & Heth, C. D. Reinstatement of fear to an extinguished conditioned stimulus. *J. Exp. Psychol. Anim. Behav. Process.* **1**, 88–96 (1975).
- Bouton, M. E. & King, D. A. Contextual control of the extinction of conditioned fear: tests for the associative value of the context. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 248–265 (1983).

6. Thomas, D. R. & Sherman, L. An assessment of the role of handling cues in "spontaneous recovery" after extinction. *J. Exp. Anal. Behav.* **46**, 305–314 (1986).
7. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454**, 600–606 (2008).
8. Burgos-Robles, A., Vidal-Gonzalez, I. & Quirk, G. J. Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *J. Neurosci.* **29**, 8474–8482 (2009).
9. Milad, M. R. & Quirk, G. J. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**, 70–74 (2002).
10. Tronson, N. C. *et al.* Segregated populations of hippocampal principal CA1 neurons mediating conditioning and extinction of contextual fear. *J. Neurosci.* **29**, 3387–3394 (2009).
11. Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
12. Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).
13. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
14. Kim, J. H. & Richardson, R. The effect of temporary amygdala inactivation on extinction and reextinction of fear in the developing rat: unlearning as a potential mechanism for extinction early in development. *J. Neurosci.* **28**, 1282–1290 (2008).
15. Hong, I., Song, B., Lee, S., Kim, J. & Choi, S. Extinction of cued fear memory involves a distinct form of depotentiation at cortical input synapses onto the lateral amygdala. *Eur. J. Neurosci.* **30**, 2089–2099 (2009).
16. Lin, C. H., Yeh, S. H., Lu, H. Y. & Gean, P. W. The similarities and diversities of signal pathways leading to consolidation of conditioning and consolidation of extinction of fear memory. *J. Neurosci.* **23**, 8310–8317 (2003).
17. Merlo, E. & Romano, A. Memory extinction entails the inhibition of the transcription factor NF- κ B. *PLoS ONE* **3**, e3687 (2008).
18. Clem, R. L. & Huganir, R. L. Calcium-permeable AMPA receptor dynamics mediate fear memory erasure. *Science* **330**, 1108–1112 (2010).
19. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
20. Condé, F., Maire-Lepoivre, E., Audinat, E. & Crepel, F. Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. *J. Comp. Neurol.* **352**, 567–593 (1995).
21. McDonald, A. J. Organization of amygdaloid projections to the mediodorsal thalamus and prefrontal cortex: a fluorescence retrograde transport study in the rat. *J. Comp. Neurol.* **262**, 46–58 (1987).
22. Sacchetti, B., Baldi, E., Lorenzini, C. A. & Bucherelli, C. Role of the neocortex in consolidation of fear conditioning memories in rats. *Exp. Brain Res.* **152**, 323–328 (2003).
23. Garcia, R., Vouimba, R. M., Baudry, M. & Thompson, R. F. The amygdala modulates prefrontal cortex activity relative to conditioned fear. *Nature* **402**, 294–296 (1999).
24. Ziv, N. E. & Smith, S. J. Evidence for a role of dendritic filopodia in synaptogenesis and spine formation. *Neuron* **17**, 91–102 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. LeDoux as well as all the members in the Gan laboratory for comments on the manuscript. This work was supported by National Institutes of Health grant NS047325 and the Investigator-Initiated Research Grant from the Alzheimer's Association (W.-B.G.), and by the National Science Foundation (#IOS-0757780), a 2008 NARSAD Independent Investigator Award and the G. Harold & Leila Y. Mathers Foundation (T.F.F.).

Author Contributions C.S.W.L. and W.-B.G. designed the imaging experiments. C.S.W.L., W.-B.G. and T.F.F. designed the behavioural paradigms. C.S.W.L. performed all the experiments and data analysis. T.F.F. helped with the behavioural data analysis. W.-B.G. supervised the work. W.-B.G., C.S.W.L. and T.F.F. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.-B.G. (gan@saturn.med.nyu.edu).

METHODS

Animals. C57BL/6 mice expressing YFP in layer-V pyramidal neurons (H-line) were purchased from the Jackson Laboratory and group-housed in the Skirball animal facility. One-month-old ($P30 \pm 1$) male mice were used in the experiments. All experiments were approved and performed in accordance with institutional guidelines.

Fear conditioning and extinction. Apparatus. Mice were trained and tested using the FreezeFrame system (Coulbourn Instruments). For training, mouse test cages equipped with stainless-steel shocking grids were connected to a precision feedback current-regulated shocker (Coulbourn Instruments). For testing, the shocking grids were replaced with non-shocking test grids that differed in texture from the shocking grids used during conditioning. Each test cage was contained in a sound-attenuating enclosure (Coulbourn Instruments). Behaviour was recorded using low-light video cameras. Stimulus presentation was automated using Actimetrics FreezeFrame software (version 2.2; Coulbourn Instruments). All equipment was thoroughly cleaned with detergent followed by water between sessions.

Fear conditioning. Mice were habituated for 2 min on a shocking grid (cage set-up A: shocking floor grids, ethanol scent). Fear conditioning was conducted with three pairings of a 30-s, 400-Hz, 80-dB auditory cue (tone A = CS1) co-terminating with a 2-s, 0.5-mA scrambled footshock (US). The intertrial interval was 15 s. Two minutes after conditioning, mice were returned to their home cages. For the unpaired control group, mice received tones and shocks in an unpaired manner (tones and shocks were separated by random intervals of 5–15 s). Mice were returned to their home cages 2 min after presentation of the unpaired stimuli. For tone-only or shock-only control groups, mice were habituated for the same amount of time in the testing context, and tones or footshocks were given separately.

Recall test and extinction. For the recall test, mice were placed in a different context (cage set-up B: test floor grids, 1% Pinesol) for an initial 2-min (pre-CS) period and this was followed by tone presentation for 2 min (CS). For extinction training, mice were subjected to five CS presentations (each lasting 2 min with an intertrial interval of 2 min) per day for two consecutive days. For repeated exposure of mice to tone B (CS2), the extinction protocol was used with a different auditory cue (1,200 Hz, 80 dB).

Reconditioning. Mice were reconditioned in the training context using shocking grids (ethanol scent) with five pairings of CS-US (30-s CS/tone A, 400 Hz; each co-terminated with a 2-s footshock) each day on days 5 and 6. For the unpaired control, five shocks and tones were given to mice in an unpaired manner each day. For conditioning with tone B, five CS-US pairings were given to the mice using tone B as the auditory cue each day.

Retrograde and anterograde tracing with Mini-Ruby. One-month-old male mice expressing YFP were anaesthetized with ketamine and xylazine (intraperitoneal; 20 mg ml⁻¹ and 3 mg ml⁻¹, respectively, in saline; 6 µl per gram of body weight). Mini-Ruby (Invitrogen) dissolved at 5% concentration in water was injected into the frontal association cortex (+2.8 mm bregma, +1.0 mm midline, +0.01 mm ventral) through a sharp electrode by iontophoresis (6 µA, on-off for 15 min). For amygdala injections, 0.15 µl of Mini-Ruby was injected over 5 min into the amygdala with a Hamilton syringe (−1.94 mm bregma, +3.00 mm midline, +4.75 mm ventral). Four to seven days after the injection, mice were perfused with 4% paraformaldehyde and their brains were postfixed overnight. Brains were sectioned with a vibratome at 200 µm. Confocal images were acquired using a Bio-Rad confocal microscope (×20 oil lens; numerical aperture, 0.8).

Transient inactivation of the frontal association cortex with muscimol. After fear conditioning or extinction, mice were anaesthetized with ketamine and xylazine. Muscimol (Sigma; 0.5 µl at 1 µg µl⁻¹) in artificial cerebrospinal fluid or artificial cerebrospinal fluid as vehicle was microinjected bilaterally into the frontal association cortex (+2.8 mm bregma, ±1.0 mm midline, +0.5 mm ventral) with a pressure injection device (Picospritzer III; 15 p.s.i., 12 ms, 0.8 Hz) over 5 min. The injection was performed within 1 h after fear conditioning or extinction. Twenty-four hours after injection, mice were subjected to a recall test. Muscimol spread was estimated by the line at which the Mini-Ruby fluorescence was less than 20% of its peak level. On the basis of this definition, we determined the range of muscimol spread as ~500–700 µm (Supplementary Fig. 3).

In vivo transcranial two-photon imaging. Spine formation and elimination were examined by imaging the mouse cortex through a thinned-skull window as described previously^{11,12}. Briefly, one-month-old male mice expressing YFP were anaesthetized with ketamine and xylazine (intraperitoneal; 20 mg ml⁻¹ and 3 mg ml⁻¹, respectively, in saline; 6 µl per gram of body weight). Thinned-skull windows were made in head-fixed mice with high-speed microdrills. Skull thickness was reduced to about 20 µm. A two-photon microscope tuned to 920 nm (×60 water immersion lens; numerical aperture, 1.1) was used to acquire images. For re-imaging of the same region, thinned regions were identified on the basis of the maps of the brain vasculature. Microsurgical blades were used to re-thin the region of interest until a clear image could be obtained. The area of the imaging region was 200 µm × 200 µm. The centres of the imaging regions were as follows: +2.8 mm bregma, +1.0 mm midline (frontal association cortex); −1.1 mm bregma, +3.4 mm midline (barrel cortex). Because repeated thinning of the skull to ~20 µm without damaging the cortex becomes more difficult after multiple imaging sessions, we designed our experiments in such a way that no animal was imaged than four times.

Data analysis. All data analysis was performed blind to treatment conditions, except for the data in Supplementary Fig. 4, which were collected under behavioural treatment conditions known to the investigator. Image J software was used to analyse spine elimination and formation from three-dimensional image stacks as described previously^{11,12}. Dendritic branches were randomly sampled within a 200 µm × 200 µm area imaged at a distance of 0–100 µm below the pia surface. The same dendritic segments were identified from three-dimensional stacks taken at different time points with high image quality (ratio of signal to background noise, >4:1). The number and location of dendritic protrusions (protrusion lengths were more than one-third of the dendritic shaft diameter) were identified. The percentage of spine formation and elimination is presented as the number of spines formed or eliminated between the first and second view divided by the total number of spines observed at the first view^{11,12} (Supplementary Information, section 2).

For the analysis of spine changes at the individual branch level (Fig. 3), the percentage of spine formation and elimination is presented as the number of formed or eliminated spines divided by the total number of spines on the individual branch. For the spine orientation analysis, newly formed spines were considered to have the same orientation when they were oriented within 90° relative to previously eliminated spines.

Filopodia were identified as long, thin structures^{11,12} (generally larger than twice the average spine length; ratio of head diameter to neck diameter, <1.2:1; ratio of length to neck diameter, >3:1). The remaining protrusions were classified as spines. No subtypes of spines were distinguished in our analysis. Three-dimensional stacks were used to ensure that tissue movements and rotation between imaging intervals did not hinder spine identification. Spines or filopodia were considered to be identical between views if their positions were unchanged with respect to adjacent landmarks. Our quantitative analysis shows that we can measure the distance between two adjacent stable spines with a precision of ~0.2 µm in 95% of the cases (2 s.d.). Spines were considered different if their positions differed from the first view by more than 0.7 µm. We chose 0.7 µm as a cut-off distance because spine positions can shift by up to ~0.3 µm in either direction along the axis of dendritic shafts owing to changes in spine morphology, slight tissue rotation and movements related to brain pulsation. We estimate that the imaging resolution of our two-photon microscope (×60; numerical aperture, 1.1; 920 nm) is ~0.7 µm.

For image display, fluorescent structures near and out of the focal plane of the dendrites of interest were removed manually from image stacks using Adobe Photoshop. The modified image stacks were then projected to generate two-dimensional images and adjusted for contrast and brightness.

For statistical analysis, we used either analysis of variance or Student's *t*-test to compare spine formation and elimination rates among different experimental groups. All spine formation and elimination rates are presented as mean ± s.d. and all behavioural data are presented as mean ± s.e.m. Data in spine distribution graphs in Fig. 3g and 4d are presented as mean ± s.e.m. The Pearson correlation coefficient was used as a measure of the strength of linear dependence between spine changes and behavioural responses. The linear regression lines and correlation coefficients (*r*) are shown in Fig. 1i–l, Fig. 2d, e, g and Fig. 3a–c. In all analyses, *P* values less than 0.05 were considered to be statistically significant.

Re-emerging superconductivity at 48 kelvin in iron chalcogenides

Liling Sun^{1*}, Xiao-Jia Chen^{2,3*}, Jing Guo¹, Peiwen Gao¹, Qing-Zhen Huang⁴, Hangdong Wang⁵, Minghu Fang⁵, Xiaolong Chen¹, Genfu Chen¹, Qi Wu¹, Chao Zhang¹, Dachun Gu¹, Xiaoli Dong¹, Lin Wang⁶, Ke Yang⁷, Aiguo Li⁷, Xi Dai¹, Ho-kwang Mao² & Zhongxian Zhao¹

Pressure has an essential role in the production¹ and control^{2,3} of superconductivity in iron-based superconductors. Substitution of a large cation by a smaller rare-earth ion to simulate the pressure effect has raised the superconducting transition temperature T_c to a record high of 55 K in these materials^{4,5}. In the same way as T_c exhibits a bell-shaped curve of dependence on chemical doping, pressure-tuned T_c typically drops monotonically after passing the optimal pressure^{1–3}. Here we report that in the superconducting iron chalcogenides, a second superconducting phase suddenly re-emerges above 11.5 GPa, after the T_c drops from the first maximum of 32 K at 1 GPa. The T_c of the re-emerging superconducting phase is considerably higher than the first maximum, reaching 48.0–48.7 K for $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ and $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$.

The recent discoveries of superconductivity at 30–32 K in a new family of iron-based chalcogenide superconductors^{6–9} $\text{A}_{1-x}\text{Fe}_2-y\text{Se}_2$ (where A = K, Rb or Cs, with possible Tl substitution) bring new excitement to the field of superconductivity¹⁰. These superconductors have unusually large magnetic moments up to $3.3\mu_B$ per Fe atom and a Fe-vacancy ordering in the Fe square lattice¹¹. How superconductivity with such a high T_c can exist on such a strong magnetic background remains perplexing¹⁰. It has been established that superconductivity in strongly correlated electronic systems can be dictated by their crystallographic structure, electronic charge, and orbital and spin degrees of freedom, which can all be manipulated by controlling parameters such as pressure, magnetic field and chemical composition^{12–15}. Pressure is a ‘clean’ way to tune basic electronic and structural properties without changing the chemistry. High-pressure studies are thus very useful in elucidating mechanisms of superconductivity as well as in searching for new high- T_c superconducting materials.

We studied single crystals of $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ and $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$ grown by the Bridgman method^{6,16,17}. We conducted both high-pressure resistance and susceptibility measurements to detect superconductivity *in situ* at high pressures and low temperatures. Figure 1 shows the temperature dependence of the electrical resistance at different pressures for $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$ single crystals. Here we define T_c as the intersection of the tangent through the inflection point of the resistive transition with a straight-line fit of the normal state just above the transition. As can be seen, T_c starts at the maximum of 33 K at 1.6 GPa, shifts to lower temperatures at increasing pressures, and vanishes near 9 GPa in our experimental temperature range, which is 300–4 K for our high-pressure resistance measurements (Fig. 1a). At slightly higher pressures, however, an unexpected superconducting phase re-emerges with an onset T_c as high as 48.0 K at 12.4 GPa (Fig. 1b). The sample is not superconducting at pressures higher than 13.2 GPa. We repeated the measurements with new samples in three independent experiments, and the results were reproducible.

To confirm the pressure-induced changes of superconductivity in $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, we also performed magnetic alternating-current susceptibility measurements *in situ* at high pressures (Fig. 2). The value of T_c is taken to be the onset of superconductivity defined by the intersection of a line drawn through the steep slope of the curve and the region of zero slope above the transition. The magnetic study showed that T_c decreased with increasing pressure and vanished at 9.8 GPa in the first superconducting phase SC-I (Fig. 2a). With further increasing pressure, the material enters a new superconducting phase SC-II and its transition temperature reaches 40.2 K at 12.2 GPa (Fig. 2b). The magnetic measurements yield T_c values consistent with the resistivity data within the experimental uncertainties. These results provide convincing evidence for the existence of two distinct superconducting phases in $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$.

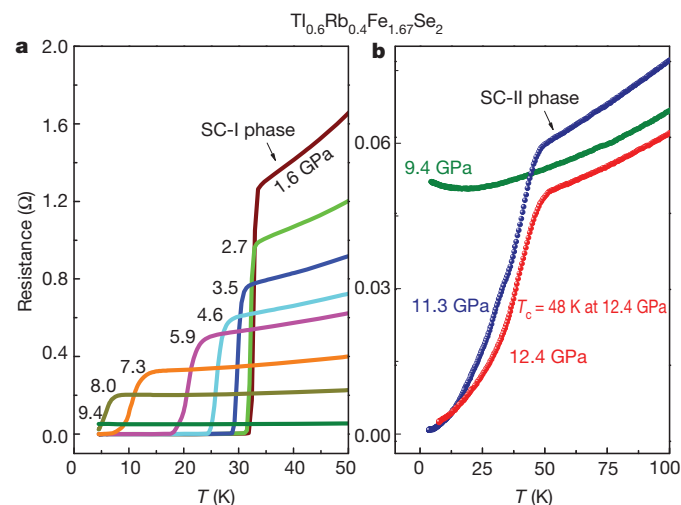


Figure 1 | Temperature dependence of electrical resistance for $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$ at different pressures. **a**, Resistance–temperature curves in the initial superconducting phase (SC-I) up to 9.4 GPa. T_c was observed to shift to lower temperature with increasing pressure. Superconductivity disappears at 9.4 GPa. **b**, Electrical resistance curves for the same single crystal at higher pressures. A new superconducting state re-emerges upon further compression. The pressure-induced superconducting phase (SC-II) has a T_c of 48 K, which is much higher than the maximum in SC-I. Cryogenic resistance measurements were performed in a diamond-anvil cell. Diamond anvils with 600- μm and 300- μm tip flats were used with sample chambers of diameter 300 μm and 100 μm , respectively. Four electrical leads were attached to the single-crystal sample insulated from the rhenium gasket, and loaded into the sample chamber. NaCl powders were employed as a pressure medium. The ruby fluorescence method was used to gauge pressure²⁰.

¹Institute of Physics and Beijing National Laboratory for Condensed Matter Physics, Chinese Academy of Sciences, Beijing 100190, China. ²Geophysical Laboratory, Carnegie Institution of Washington, Washington DC 20015, USA. ³Department of Physics, South China University of Technology, Guangzhou 510640, China. ⁴NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA. ⁵Department of Physics, Zhejiang University, Hangzhou 310027, China. ⁶HPSynC, Geophysical Laboratory, Carnegie Institution of Washington, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. ⁷Shanghai Synchrotron Radiation Facilities, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201204, China.

*These authors contributed equally to this work.

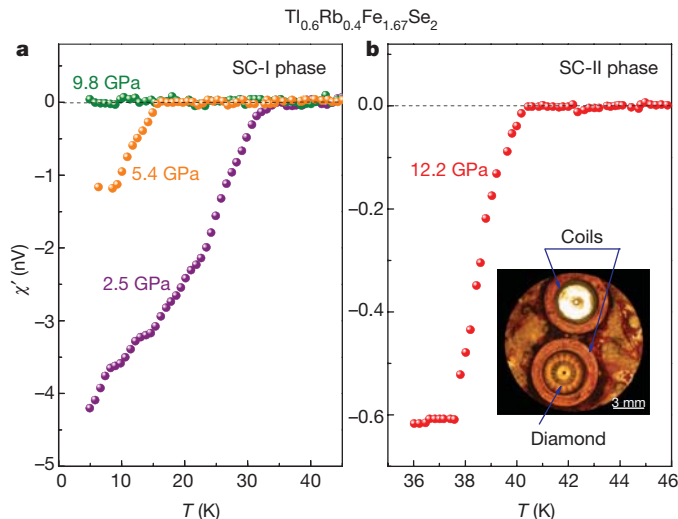


Figure 2 | Temperature dependence of the alternating-current susceptibility for $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$ at different pressures. **a**, Superconducting transitions observed in the real susceptibility component of the sample at pressures of 2.5, 5.4 and 9.8 GPa in SC-I. The superconducting transition shifts downward to lower temperature with increasing pressure. At 9.8 GPa the susceptibility component remains constant upon cooling down to 4 K, indicating that the sample is no longer superconducting. **b**, The real component of the susceptibility versus temperature for the crystal in SC-II at a pressure of 12.2 GPa. The inset shows the set-up for alternating-current susceptibility measurements in a diamond-anvil cell, with a signal coil around the diamond anvils and a compensating coil. The alternating-current susceptibilities were detected within a lock-in amplifier²¹. The crystals were loaded into the sample chamber, which is a hole in the centre of the nonmagnetic gaskets, with Daphne 7373 as the pressure medium.

To investigate whether the pressure-induced re-emergence of superconductivity was unique to $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$ or more general among iron chalcogenides, we conducted parallel electrical resistance measurements on $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ single crystals, and observed nearly identical behaviour (Fig. 3). The initial T_c of 32 K at 0.8–1.6 GPa decreased monotonically with increasing pressure and became undetectable at 9.2 GPa. At a slightly increased pressure, the second superconducting phase of $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ re-emerged and reached the maximum T_c of 48.7 K

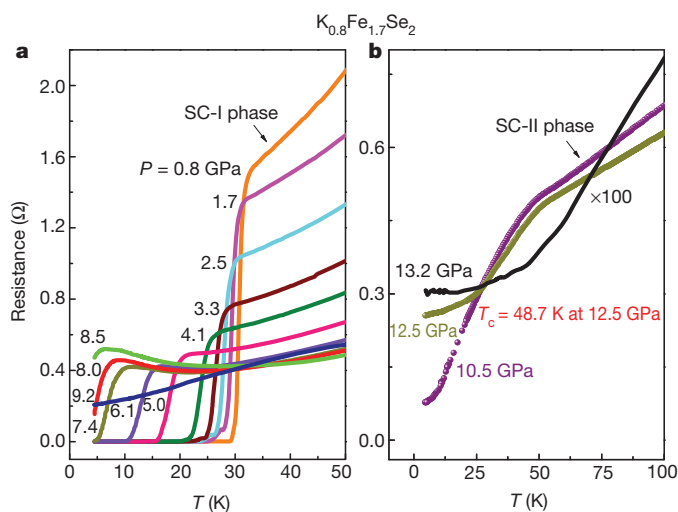


Figure 3 | Temperature dependence of the resistance for $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ at different pressures. **a**, SC-I. The resistance–temperature curves showing the T_c reduction with increasing pressure and its disappearance at 9.2 GPa. **b**, SC-II. The resistance measurements reveal another superconducting phase above 10.5 GPa. The T_c reaches 48.7 K at 12.5 GPa and disappears at 13.2 GPa. The black curve has been multiplied by 100.

at 12.5 GPa. We repeated the experiment six times using six single crystals cut from different batches, and the results were reproducible. We further repeated the measurements with a slightly different composition, $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$, and again, observed similar pressure-induced behaviour.

We summarized the pressure dependence of T_c of $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$, and $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$ in Fig. 4 and Supplementary Tables 1 to 4. The diagram clearly reveals two distinct superconducting regions: the initial superconducting phase SC-I and the pressure-induced superconducting phase SC-II. In the SC-I region, T_c is suppressed with applied pressure and approaches zero between 9.2 and 9.8 GPa. At higher pressures, the SC-II region appears, in which the T_c is even higher than the maximum T_c of the SC-I region. The SC-II region has a maximum T_c of 48.7 K for $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ and 48.0 K for $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, higher than previously observed in chalcogenide superconductors. The SC-II region appears in a narrow pressure range. Unlike the usual parabolic pressure-tuning curve of T_c , the high T_c in SC-II appears abruptly above 9.8 GPa and disappears equally abruptly above 13.2 GPa. Intermediate $T_c < 38$ K is not observed even with small pressure increment steps of 0.1 GPa. A similar re-emergence of superconductivity has been observed in some other strongly correlated electronic systems, such as heavy-fermion¹² and organic systems¹⁸.

Our preliminary high-pressure polycrystalline X-ray diffraction results of the two iron chalcogenides $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ and $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$ confirm that, to the first degree, the basic tetragonal crystal structure persists throughout the pressure range studied (Supplementary Information). Therefore, the disappearance of T_c in SC-I, the re-emergence of higher T_c in SC-II, and the final non-superconducting region reflect detailed structural variances within the basic tetragonal unit cell, which await future in-depth investigation with advanced diagnostic probes. For instance, the possible change in magnetic-ordering structures would require high-pressure neutron diffraction, and the possible superlattice and Fe vacancy ordering would require high-pressure single-crystal X-ray structural investigations.

The pressure dependence of T_c in the SC-I region is expected but its mechanism is still much debated. Quantum criticalities are thought to affect superconductivity for strongly correlated electronic systems¹⁹. A characteristic feature of the new iron chalcogenide superconductors is

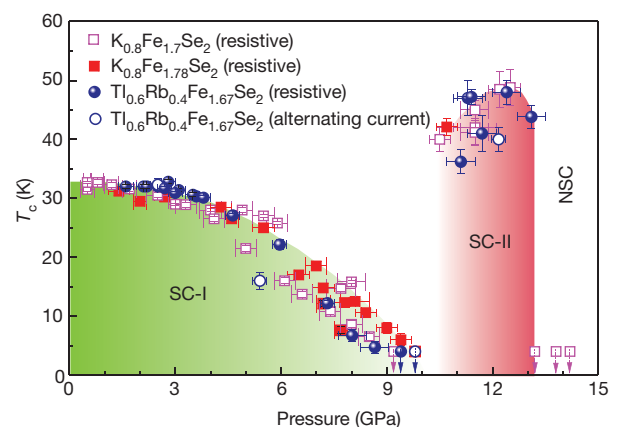


Figure 4 | Pressure dependence of the T_c for $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ and $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$. The symbols represent the pressure–temperature conditions for which T_c values were observed from the resistive and alternating-current susceptibility measurements; symbols with downward arrows represent the absence of superconductivity to the lowest temperature (4 K). All $\text{Tl}_{0.6}\text{Rb}_{0.4}\text{Fe}_{1.67}\text{Se}_2$, $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ and $\text{K}_{0.8}\text{Fe}_{1.78}\text{Se}_2$ samples show two superconducting regions (SC-I and SC-II) separated by a critical pressure at around 10 GPa. NSC, the non-superconducting region above 13.2 GPa. The maximum T_c is found to be 48.7 K in $\text{K}_{0.8}\text{Fe}_{1.7}\text{Se}_2$ at a pressure of 12.5 GPa. At higher pressures above 13.2 GPa, the samples are non-superconducting. Error bars are one standard deviation.

the existence of Fe-vacancies in the Fe-square lattice, ordered by a $\sqrt{5} \times \sqrt{5}$ superstructure¹¹. It remains unclear whether pressure could destroy the vacancy ordering at a critical value and drive the materials into a disordered lattice. Detailed structural studies of these superconducting behaviours in the iron chalcogenide superconductors are currently being conducted. Their magnetic properties at high pressures should help us to understand the interplay of magnetism and superconductivity in these iron chalcogenides.

This observation of the SC-II region with the re-emerging higher T_c is unexpected. It will certainly stimulate a great deal of future experimental and theoretical studies to clarify whether the observed re-emergence of superconductivity in iron chalcogenides is associated with the quantum critical transition, magnetism, superstructure, vacancy ordering or spin fluctuation.

Received 21 October; accepted 19 December 2011.

Published online 22 February 2012.

1. Torikachvili, M. S., Bud'ko, S. L., Ni, N. & Canfield, P. C. Pressure induced superconductivity in CaFe_2As_2 . *Phys. Rev. Lett.* **101**, 057006 (2008).
2. Takahashi, H. *et al.* Superconductivity at 43 K in an iron-based layered compound $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$. *Nature* **453**, 376–378 (2008).
3. Medvedev, S. *et al.* Electronic and magnetic phase diagram of $\beta\text{-Fe}_{1.01}\text{Se}$ with superconductivity at 36.7 K under pressure. *Nature Mater.* **8**, 630–633 (2009).
4. Chen, X. H. *et al.* Superconductivity at 43 K in $\text{SmFeAsO}_{1-x}\text{F}_x$. *Nature* **453**, 761–762 (2008).
5. Ren, Z. A. *et al.* Superconductivity at 55 K in iron-based F-doped layered quaternary compound $\text{Sm}[\text{O}_{1-x}\text{F}_x]\text{FeAs}$. *Chin. Phys. Lett.* **25**, 2215–2216 (2008).
6. Guo, J. G. *et al.* Superconductivity in the iron selenide $\text{K}_x\text{Fe}_2\text{Se}_2$ ($0 < x < 1.0$). *Phys. Rev. B* **82**, 180520(R) (2010).
7. Krzton-Maziopa, A. *et al.* Synthesis and crystal growth of $\text{Cs}_{0.8}(\text{FeSe}_{0.98})_2$: a new iron-based superconductor with $T_c = 27$ K. *J. Phys. Condens. Matter* **23**, 052203 (2011).
8. Wang, A. F. *et al.* Superconductivity at 32 K in single-crystalline $\text{Rb}_x\text{Fe}_{2-y}\text{Se}_2$. *Phys. Rev. B* **83**, 060512(R) (2011).
9. Fang, M. H. *et al.* Fe-based superconductivity with $T_c = 31$ K bordering an antiferromagnetic insulator in $(\text{Ti},\text{K})\text{Fe}_x\text{Se}_2$. *Europhys. Lett.* **94**, 27009 (2011).
10. Mazin, I. Iron superconductivity weathers another storm. *Physics* **4**, 26 (2011).
11. Bao, W. *et al.* A novel large moment antiferromagnetic order in $\text{K}_{0.8}\text{Fe}_{1.6}\text{Se}_2$ superconductor. *Chin. Phys. Lett.* **28**, 086104 (2011).
12. Yuan, H. Q. *et al.* Observation of two distinct superconducting phases in CeCu_2Si_2 . *Science* **302**, 2104–2107 (2003).
13. Chen, X. J. *et al.* Enhancement of superconductivity by pressure-driven competition in electronic order. *Nature* **466**, 950–953 (2010).
14. Uji, S. *et al.* Magnetic-field-induced superconductivity in a two-dimensional organic conductor. *Nature* **410**, 908–910 (2001).
15. Jin, K., Butch, N. P., Kirshenbaum, K. & Greene, R. L. Link between spin fluctuations and electron pairing in copper oxide superconductors. *Nature* **476**, 73–75 (2011).
16. Wang, D. M., He, J. B., Xia, T.-L. & Chen, G. F. The effect of varying Fe-content on transport properties of K intercalated iron selenide $\text{K}_x\text{Fe}_{2-y}\text{Se}_2$. *Phys. Rev. B* **83**, 132502 (2011).
17. Wang, H. D. *et al.* Superconductivity at 32 K and anisotropy in $\text{Tl}_{0.58}\text{Rb}_{0.42}\text{Fe}_{1.72}\text{Se}_2$ crystals. *Europhys. Lett.* **93**, 47004 (2011).
18. Okuhata, T. *et al.* High-pressure studies of doped-type organic superconductors. *J. Phys. Soc. Jpn* **76** (Suppl. A) 188–189 (2007).
19. Si, Q. M. & Steglich, F. Heavy fermions and quantum phase transitions. *Science* **329**, 1161–1166 (2010).
20. Mao, H. K., Xu, J. & Bell, P. M. Calibration of the ruby pressure gauge to 800 Kbar under quasi-hydrostatic conditions. *J. Geophys. Res.* **91**, 4673–4676 (1986).
21. Debessai, M., Matsuoka, T., Hamlin, J. J. & Schilling, J. S. Pressure-induced superconducting state of europium metal at low temperatures. *Phys. Rev. Lett.* **102**, 197002 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. I. Mazin, W. Bao, and T. Xiang for discussions and J. S. Schilling for the help with the alternating-current susceptibility technique. This work in China was supported by the NSCF, 973 projects, and Chinese Academy of Sciences. This work in the USA was supported as part of the EFree, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences (DOE-BES). The High Pressure Collaborative Access Team (HPCAT) is supported by CIW, CDAC, UNLV and LLNL through funding from the DOE-NNSA, the DOE-BES and the NSF. The Advanced Photon Source (APS) is supported by the DOE-BES.

Author Contributions L.S., X.-J.C., H.M. and Z.Z. designed the project; X.-J.C., L.S., H.M., Q.W. and Z.Z. wrote the paper; P.G., J.G., C.Z. and L.S. performed the resistance and magnetic susceptibility measurements; X.-J.C., L.W., J.G., P.G., L.S., K.Y. and A.L. performed synchrotron X-ray diffraction measurements; and M.F., X.C. and G.C. synthesized the single crystals. All the authors analysed the data and discussed the results. All the authors read and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Z.Z. (zhxzhao@iphy.ac.cn) or H.M. (hmao@gl.ciw.edu).

Surprisingly complex community discovered in the mid-Devonian fossil forest at Gilboa

William E. Stein¹, Christopher M. Berry², Linda VanAller Hernick³ & Frank Mannolini³

The origin of trees by the mid-Devonian epoch (398–385 million years ago) signals a major change in terrestrial ecosystems with potential long-term consequences including increased weathering, drop in atmospheric CO₂, modified climate, changes in sedimentation patterns and mass extinction^{1–3}. However, little is known about the ecology of early forests or how changes in early terrestrial ecosystems influenced global processes. One of the most famous palaeontological records for this time is the ‘oldest fossil forest’ at Riverside Quarry, Gilboa, New York, USA, discovered in the 1920s^{4,5}. Hundreds of large *Eospermatopteris* sandstone casts, now thought to represent the bases of standing cladoxylopsid trees⁶, were recovered from a horizon that was originally interpreted as a muddy swamp. After quarry operations ceased, relatively minor outcrops of similar fossils at nearby localities have provided limited opportunities to evaluate this pervasive view using modern methods^{7,8}. In 2010, removal of the quarry backfill enabled reappraisal of the palaeoecology of this important site. Here we describe a 1,200 m² map showing numerous *Eospermatopteris* root systems in life position within a mixed-age stand of trees. Unexpectedly, large woody rhizomes with adventitious roots and aerial branch systems identified as aneurophytalean progymnosperms run between, and probably climb into, *Eospermatopteris* trees. We describe the overall habit for these surprisingly large aneurophytaleans, the earliest fossil group having wood produced by a bifacial vascular cambium. The site also provides evidence for arborescence within lycopsids, extending the North American range for trees in this ecologically critical group. The rooting horizon is a dark grey sandy mudstone showing limited root penetration. Although clearly belonging to a wetland coastal plain environment⁹, the forest was probably limited in duration and subject to periodic disturbance. These observations provide fundamental clarification of the palaeoecology of this mixed-group early forest, with important implications for interpreting coeval assemblage data worldwide.

Tree-sized plants become evident in the fossil record in the Middle Devonian. The oldest (primarily from the Eifelian–Givetian ages) belong to the order Pseudosporochnales (class Cladoxylopsida). These plants possess a strong central trunk bearing a crown of ephemeral non-leafy branches and have an overall form that looks much like a modern palm or tree-fern¹⁰. Recent evidence unites the long-enigmatic *Eospermatopteris* stump casts from Riverside Quarry, Gilboa (Givetian age) with pseudosporochnalean crown and branch compressions of the genus *Wattieza*⁶. Although the group has almost worldwide distribution, little is currently known about its morphological variation, phylogenetic relationships or occurrence in specific palaeoenvironments. As a result, Goldring’s iconic “oldest fossil forest” (ref. 5) continues to exert great influence. She described trees “resting in and upon shale” representing “muds in which the trees stood”, with bases that were “bulbous, as might be expected of certain trees growing under swampy conditions”⁵. By the late Middle Devonian (Givetian), pseudosporochnaleans were joined by the order Archaeopteridales (class Progymnospermopsida)¹¹.

Demonstration of a bifacial vascular cambium producing wood, leaves, extended development by means of buds¹², and possibly more deeply penetrating roots, has led to the hypothesis that archaeopteridaleans mark the beginning of another type of forest in the Devonian that is different from the water-restricted *Eospermatopteris*^{11,13}. However, both groups are commonly found together in variable lithologies within the North American Catskill sediment complex¹⁴ and elsewhere, with little to suggest spatial or environmental separation of the plants.

In spring 2010, backfill at Riverside Quarry was removed, exposing the highwall and extensive quarry floor at the original 292.6-m (960-ft) elevation for a few weeks, before refilling. The highwall section consists of several massive fine sandstone beds that are 0.2–0.8 m thick, with one thin wedge of shale (Supplementary Figs 1 and 2). Two sandstone moulds left by *Eospermatopteris* casts collected in the 1920s occur at the 296-m (972-ft) level⁴. Both moulds show attached roots forming a loose root mantle within the enclosing sandstone, passing outwards, downwards and partly penetrating a palaeosol below (Fig. 1a).

Of approximately 30 new *Eospermatopteris* casts retrieved from the backfill, two specimens remain encased within the casting sandstone with their flat to shallowly convex bottoms exposed and their roots still attached (Fig. 1b and Supplementary Fig. 8). Numerous and conspicuous raised mounds of radiating roots delimit the original, and still largely intact, forest palaeosol surface on the quarry floor (Fig. 1c and Supplementary Figs 3–11). Root mounds occur over most, if not all, of the quarry floor and contact of palaeosol with the overlying sandstone responsible for *Eospermatopteris* casts is clearly observed at the base of the highwall. We conclude that the root mounds on the quarry floor and exposed cast bottoms described above represent matching counterpart surfaces. Approximately 1,200 m² of the palaeosol surface was pressure-washed and plan-mapped, and this showed a total of 486 objects including root mounds, horizontal stems and aerial plant fragments (Fig. 2). The basic features of this survey and of the criteria that we used for recognizing different plants are described here and in the Supplementary Information.

Approximately 200 root mounds in the mapped region unquestionably mark locations of *Eospermatopteris* casts that were originally removed from the site. The best-preserved examples typically exhibit a circular densely imbricate skirt of unbranched lateral roots 1–2 cm in diameter radiating from a raised rim and central basin corresponding respectively to the widest diameter and bottom of the originally seated cast. The rim, approximately 0.2–1.9 m in diameter, suggests outward displacement of soil by each tree base during growth. The central basin often shows a radiating pattern of grooves with circular depressions, which correspond to roots descending directly into the quarry floor. Although most root mounds represent single individuals, some more extensive mounds exhibit multiple circular to elliptical basins each with a raised rim suggesting the presence of two or more intertwined individuals or genetically related ramets sharing a common root mass (Supplementary Fig. 6). One specimen (Supplementary Fig. 7) was evidently toppled and shows the base of a horizontal trunk and a root mass oriented at right angles to their typical appearance on the quarry

¹Department of Biological Sciences, Binghamton University, New York 13902-6000, USA. ²School of Earth and Ocean Sciences, Cardiff University, Cardiff CF10 3YE, UK. ³New York State Museum, Albany, New York 12230, USA.

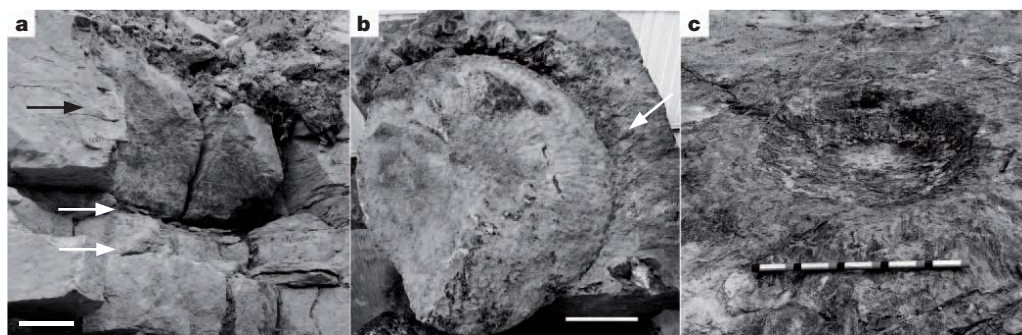


Figure 1 | *Eospermatopteris* cast, moulds and root system, Riverside Quarry, New York, USA. **a**, Mould in quarry highwall showing casting sandstone (black arrow) and the vertical extent of rooted palaeosol (white arrows) with both levels containing attached roots. Scale bar, 20 cm. **b**, Bottom of newly recovered cast with attached roots diverging into casing sandstone

(arrow; New York State Museum (NYSM) 18055). Scale bar, 20 cm. **c**, Forest palaeosol horizon on quarry floor showing raised rim, central basin and extensive rooting system; the surface equivalent to the counterpart of the cast shown in **b**. Entire scale bar, 50 cm.

floor. Heterogeneity in spacing of *Eospermatopteris* trees is evident along with a size variation consistent with individuals of different ages.

A surprising second tree-sized member of this assemblage is represented by straight to curved horizontal stems that are as much as 15 cm in diameter, probably greatly exceeding 4 m in length (Figs 2 (arrows b and h) and 3a, and Supplementary Figs 16–20). In a small number of cases, segments of stem are enveloped by numerous unbranched lateral organs that are approximately 5 mm in diameter, interpreted as roots (Fig. 3b). Surfaces of stems are unique in the quarry in having a 1–3-mm thick outer carbon layer together with an orthostichous pattern of circular to transversely elongate depressions, or corresponding knobs permineralized by pyrite. The permineralizations show fragmentary evidence of original tissues consistent with broken branch

bases oriented at right angles to the main stem preserved long after they originally functioned (Fig. 3c, d and Supplementary Figs 12–15). Secondary xylem is abundant, often interdigitating with carbon on the surface. In a transverse view of the branch bases, narrow vascular rays extend outward in radial files from secondary xylem to the vascular cambium, and beyond to sclerenchymatous elements of secondary phloem. The cortex consists of thin-walled parenchyma and clusters of sclereids. In the samples recovered so far, three- or four-ribbed protostelic configuration of the primary xylem (Fig. 3c, d and Supplementary Fig. 12) allows us to identify these within the order Aneurophytales (class Progymnospermopsida)^{15,16}. This interpretation is strongly supported by complementary compression evidence from a specimen originally collected by Goldring (Fig. 3e and Supplementary

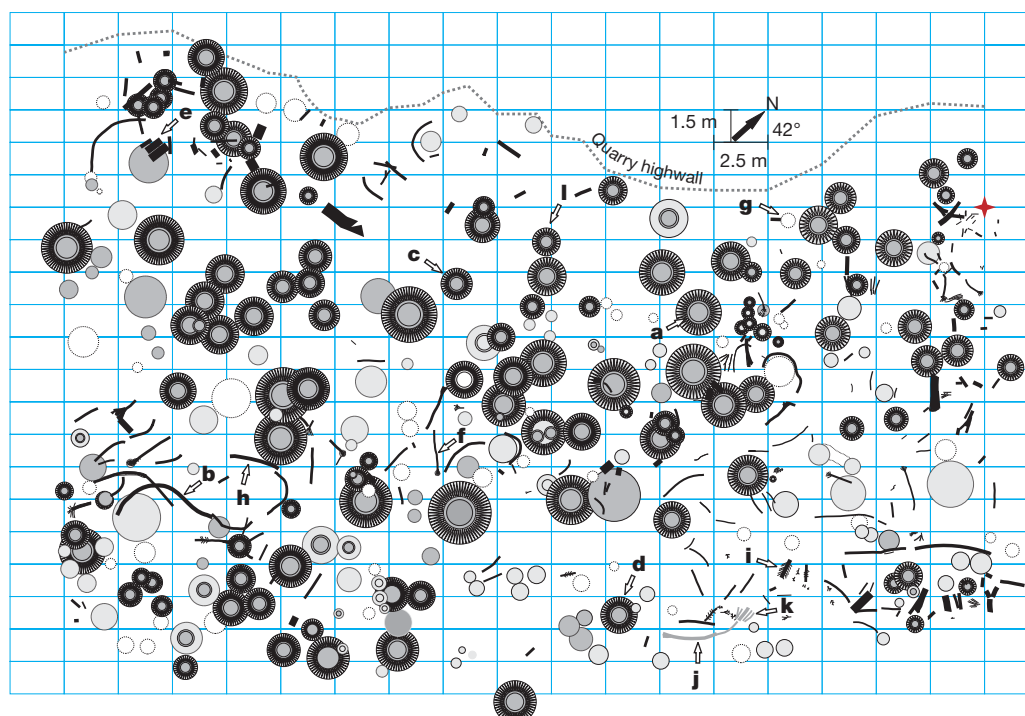


Figure 2 | Plan map of part of the quarry floor showing original rooting horizon. *Eospermatopteris* root mounds indicated with different levels of confidence. Dark shading, high confidence; light shading, intermediate confidence; dashed circles with no shading, low confidence. Radiating lines, radiating pattern of roots associated with root mound; central double circle, approximate diameter of central basin; outer circle, entire diameter of observed

root system. Aneurophytalean main stems and other linear stem fragments are shown in black. Lycopsid main stem (arrow j) and probable distal branches (arrow k) are grey. The approximate position of the quarry highwall is indicated by a dotted line; the georeference point (see Supplementary Discussion) is indicated by a four-point star. Arrows a–l, plants that are shown in Figures or in the Supplementary Information.

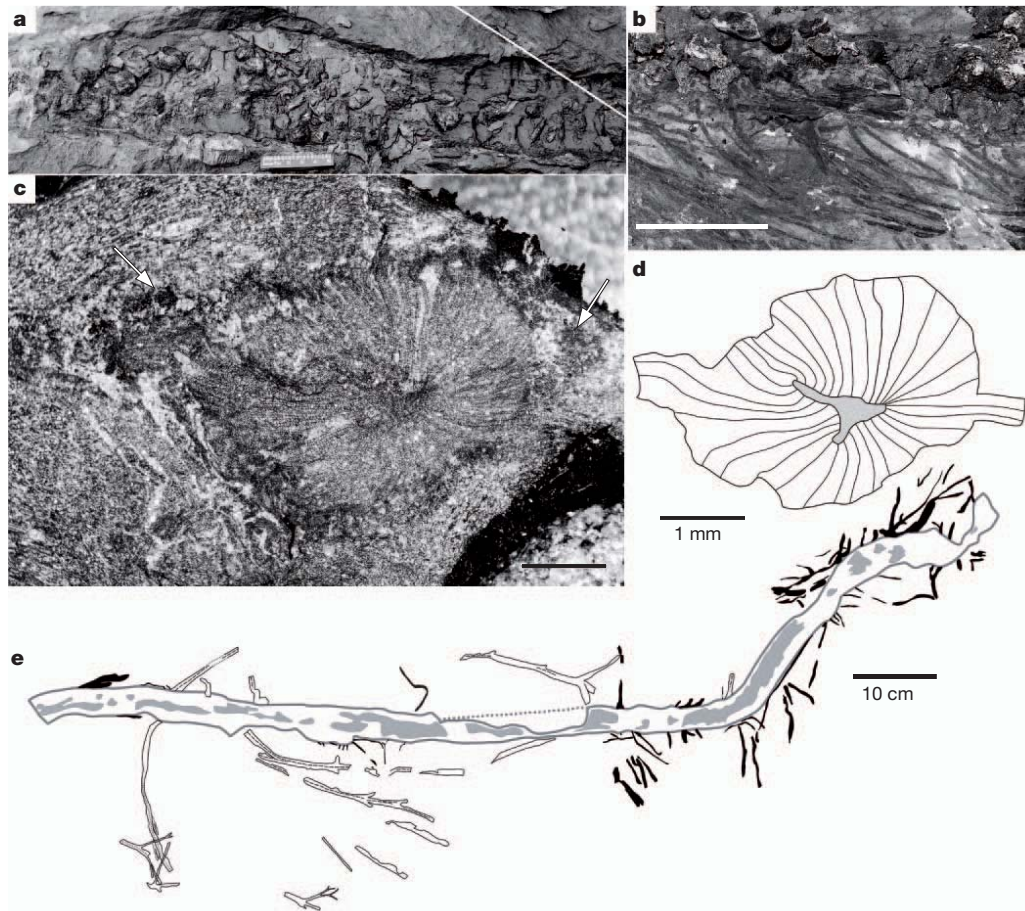


Figure 3 | Aneurophytalean rhizomes. **a**, Large specimen showing abundant carbon and several depressions, or knobs partially permineralized with pyrite. White line, part of the grid system for the map constructed in the field. Scale, 5 cm. **b**, Rhizome collected by Goldring (NYSM 6575) showing numerous attached roots. Scale bar, 10 cm. **c**, Transverse section of branch base showing primary xylem, secondary xylem, secondary phloem and remnants of cortex

(arrows), all embedded in extensive secondary xylem of the main stem (NYSM 18056). Scale bar, 1 mm. **d**, Interpretation of the lateral branch xylem in **c** showing probable three-ribbed primary xylem (shaded). **e**, Interpretation of rhizome originally collected by Goldring (NYSM 18051), showing attached and associated aerial branches in outline, attached roots in black and thick carbon on rhizome in grey.

Figs 21 and 23), also from Riverside Quarry. In this specimen a main stem shows similar orthostichous knobs bearing numerous roots of different sizes. Removal of sandstone reveals attached aerial branches that in turn bear dichotomous appendages mostly in an opposite, probably decussate, pattern that is consistent with the aneurophytalean genus *Tetraxylopteris*¹⁷.

Clustering of aneurophytaleans is conspicuous in Fig. 2, with some stems describing arcs around, and others running directly towards or away from, apparently established *Eospermatopteris* root mounds, and several stems appearing to end abruptly on an *Eospermatopteris* mound, near its rim. These findings suggest a rhizomatous habit for the plant with definite interaction between the aneurophytaleans and already established upright *Eospermatopteris* trees. It seems possible, perhaps likely, that the aneurophytaleans used the *Eospermatopteris* trees for support. Taken together, the observations suggest a new whole-plant concept for at least one member of the Aneurophytales as a large, woody, scrambling to lianaceous plant bearing adventitious roots and three-dimensional aerial branch systems functioning as frond-like ephemeral modules. These modules comprise the bulk of described evidence for this group to date^{16,17}.

At this quarry site we also found a single horizontally preserved tree clearly representing an arborescent member of the class Lycopsidea (club mosses)^{18,19} (Fig. 2, arrows j and k). An obliquely preserved main trunk that is greater than 11 cm in diameter is smooth with broad longitudinal ridges and curved over 3.9 m of preserved length (Supplementary Fig. 25). Distally, a 15-cm-diameter main trunk is

surrounded by four to six smaller axes of 4.5–8.5 cm in diameter (Fig. 4, and Supplementary Figs 26 and 27). However, the overall plant architecture is unclear. Both trunk and axes are covered with the same

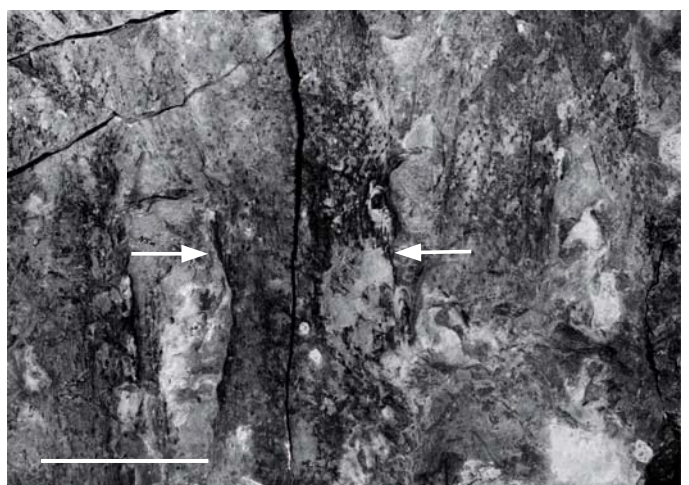


Figure 4 | Arborescent lycopsid. Upper portion of tree (arrow k in Fig. 2), including the main stem (occupying the width between the arrows) and smaller stems with similar leaf bases that probably represent distal portions of the main stem or lateral branches (NYSM 18053). Scale bar, 10 cm.

distinctive pattern of large attachment scars showing remnants of distal laminae; we can therefore confidently identify these fossils as lycopsids. Previous studies have reported abundant small lycopsids at Riverside Quarry and other Catskill localities^{20,21}, but none have been as large as this one. This record pre-dates the previous earliest (Frasnian age) lycopsid trees in North America²².

The remarkable *in situ* preservation of the Gilboa forest rooting horizon substantially enhances our understanding of this famous mid-Devonian 'earliest forest' terrestrial ecosystem. Rather than being a distinct muddy swamp containing a homogeneous stand of specially adapted *Eospermatopteris* trees, the site instead agrees sedimentologically with other exposures in the general vicinity^{23,24}, and with palaeobotanical evidence suggesting mixed assemblages worldwide²⁵. The location of this site was probably near an ancient shoreline during an interval of relatively high sea-level^{26,27}. Low-angle cross-bedded sandstones in the Gilboa region containing several *Eospermatopteris* levels suggest that the site was formed by recurrent, marine-influenced and possibly catastrophic, processes with a relatively high frequency of disturbance. Drab colour and high organic content of the forest palaeosol at Riverside Quarry indicates elevated water tables. Pyrite in aneurophytaleans, probably related to the presence of woody tissues, suggests concurrent or early diagenetic anoxia. All of this contrast with the predominantly red vertisols, which are also commonly observed in the Catskill complex. However, both facies are widespread and neither argues for a specific palaeoenvironmental reconstruction²⁸.

From one remarkable site, we now know that at least three major types of trees occurred in close proximity in a mixed, probably disturbed, wetland assemblage, with archaeopteridaleans absent at the scale of our map. We suspect that the sizes that were reached by the plants most probably relate to the rate of growth and frequency of reproduction permitted by disturbances, rather than reflect more regional environmental factors. From accumulating palaeobotanical evidence, we suggest that the very large but non-woody *Eospermatopteris* was probably hollow, reed-like and potentially fast-growing. This tree may have been capable of rapid establishment and, owing to the construction of sizeable root masses, contributed substantially to local substrate stability. Scrambling aneurophytaleans were also large, but woody, possibly longer-lived and potentially the greater contributor to global sedimentary carbon. Both plants bore leafless branch systems, suggesting only partial forest shade, and are commonly found together in Middle Devonian deposits today, probably as cast-off modular detritus. The apparent difference in growth form between rhizomatous aneurophytaleans and conifer-like archaeopteridaleans may signal some type of habitat specialization, but probably within a proximate wetland setting. Arboresecent mid-Devonian lycopsids, now clearly present in mid-Devonian early forests, do not yet seem to be associated with coal-forming environments such as are typical for the Carboniferous period. Substantially modifying the image originally provided by Goldring⁵, we now glimpse a diversified early landscape already reflecting at least some of the complexity of later terrestrial ecosystems.

Received 31 October 2011; accepted 23 December 2011.

- Algeo, T. J., Scheckler, S. E. & Maynard, J. B. in *Plants Invade the Land: Evolutionary and Environmental Approaches* (eds Gensel, P. G. & Edwards, D.) 213–236 (Columbia Univ. Press, 2001).
- Davies, N. S. & Gibling, M. R. Paleozoic vegetation and the Siluro–Devonian rise of fluvial lateral accretion sets. *Geology* **38**, 51–54 (2010).
- Driese, S. G. & Mora, C. I. in *Plants Invade the Land: Evolutionary and Environmental Approaches* (eds Gensel, P. G. & Edwards, D.) 237–253 (Columbia Univ. Press, 2001).
- Goldring, W. The Upper Devonian forest of seed ferns in eastern New York. *NY State Mus. Bull.* **251**, 50–72 (1924).
- Goldring, W. The oldest known petrified forest. *Sci. Mon.* **24**, 514–529 (1927).
- Stein, W. E., Mannolini, F., Hernick, L. V., Landing, E. & Berry, C. M. Giant cladoxylopsid trees resolve the enigma of the Earth's earliest forest stumps at Gilboa. *Nature* **446**, 904–907 (2007).
- Driese, S., Mora, C. I. & Elick, J. M. Morphology and taphonomy of root and stump casts of the earliest trees (Middle to Late Devonian), Pennsylvania and New York, U.S.A. *Palaio* **12**, 524–537 (1997).
- Mintz, J. S., Driese, S. G. & White, J. D. Environmental and ecological variability of Middle Devonian (Givetian) forests in Appalachian Basin paleosols, New York, United States. *Palaio* **25**, 85–96 (2010).
- Greb, S. F., DiMichele, W. A. & Gastaldo, R. A. Evolution and importance of wetlands in earth history. *Geol. Soc. Am. Spec. Pap.* **399**, 1–40 (2006).
- Berry, C. M. & Fairon-Demaret, M. The architecture of *Pseudosporochnus nodosus* Leclercq et Banks: a Middle Devonian Cladoxylopsid from Belgium. *Int. J. Plant Sci.* **163**, 699–713 (2002).
- Beck, C. B. Current status of the Progymnospermopsida. *Rev. Palaeobot. Palynol.* **21**, 5–23 (1976).
- Meyer-Berthaud, B., Scheckler, S. E. & Wendt, J. *Archaeopteris* is the earliest known modern tree. *Nature* **398**, 700–701 (1999).
- Retallack, G. J. & Huang, C. Ecology and evolution of Devonian trees in New York, U.S.A. *Palaeoecogr. Palaeoclimatol. Palaeoecol.* **299**, 110–128 (2011).
- Banks, H. P., Grierson, J. D. & Bonamo, P. M. The flora of the Catskill clastic wedge. *Geol. Soc. Am. Spec. Pap.* **201**, 125–141 (1985).
- Beck, C. B. & Wight, D. C. in *Origin and Evolution of Gymnosperms* (ed. Beck, C. B.) 1–84. Columbia Univ. Press, New York, 1988).
- Dannenhofer, J. M., Stein, W. & Bonamo, P. M. The primary body of *Rellimia thomsonii*: integrated perspective based on organically connected specimens. *Int. J. Plant Sci.* **168**, 491–506 (2007).
- Hammond, S. E. & Berry, C. M. A new species of *Tetraxylopteris* (Aneurophytales) from the Devonian of Venezuela. *Bot. J. Linn. Soc.* **148**, 275–303 (2005).
- Pigg, K. B. Evolution of Isoetalean lycopsids. *Ann. Mo. Bot. Gard.* **79**, 589–612 (1992).
- Gensel, P. G. & Berry, C. M. Early lycopphyte evolution. *Am. Fern J.* **91**, 74–98 (2001).
- Grierson, J. D. & Banks, H. P. Lycopsids of the Devonian of New York State. *Palaentogr. Am.* **4**, 217–295 (1963).
- Berry, C. M. & Edwards, D. A new species of the lycopsid *Gilboaphyton* Arnold from the Devonian of Venezuela and New York State with a revision of the closely related genus *Archaeosigillaria* Kidston. *Rev. Palaeobot. Palynol.* **96**, 47–70 (1997).
- White, D. A remarkable fossil tree trunk from the middle Devonian of New York. *NY State Mus. Bull.* **107**, 328–340 (1907).
- Miller, M. F. & Woodrow, D. L. in *Dynamic Stratigraphy and Depositional Environments of the Hamilton Group (Middle Devonian) in New York State, Part II* (eds Landing, E. & Brett, C. E.) 153–177 (State Education Department, 1991).
- Bridge, J. S. The geometry, flow patterns and sedimentary processes of Devonian rivers and coasts, New York and Pennsylvania, USA. *Geol. Soc. Lond. Spec. Publ.* **180**, 85–108 (2000).
- Berry, C. M. & Fairon-Demaret, M. in *Plants Invade the Land: Evolutionary and Environmental Approaches* (eds Gensel, P. G. & Edwards, D.) 120–139 (Columbia University Press, 2001).
- Bridge, J. S. & Willis, B. J. Marine transgressions and regressions recorded in Middle Devonian shore-zone deposits of the Catskill clastic wedge. *Geol. Soc. Am. Bull.* **106**, 1440–1458 (1994).
- Baird, G. C. & Brett, C. E. Late Givetian (Taghanic) bioevents in New York State: new discoveries and questions. *Bull. Geosci.* **83**, 357–375 (2008).
- Wright, V. P., Taylor, K. G. & Beck, V. H. The paleohydrology of lower Cretaceous seasonal wetlands, Isle of Wight, southern England. *J. Sediment. Res.* **70**, 619–632 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support was provided by the New York State Museum. C.M.B. acknowledges support from the UK Natural Environment Research Council (NERC) grant NE/F010699/1. Permission to work at the site by New York City Department of Environmental Protection (DEP) is acknowledged. On-site and other assistance was provided by E. Chase, P. Costa and G. Heath (DEP), as well as by P. Kollak and C. Tompkins of the Thalle Construction Company. We also thank P. Wright for advice and guidance on the major palaeosol horizon.

Author Contributions L.V.H. and F. M. were responsible for collections and analysis, W.E.S. for mapping, C.M.B. for geological section, and W.E.S. and C.M.B. for palaeoecological interpretation. W.E.S. led the writing of the paper with substantial contributions from C.M.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.E.S. (stein@binghamton.edu).

Structural basis for iron piracy by pathogenic *Neisseria*

Nicholas Noinaj¹, Nicole C. Easley¹, Muse Oke¹, Naoko Mizuno², James Gumbart³, Evzen Boura¹, Ashley N. Steere⁴, Olga Zak⁵, Philip Aisen⁵, Emad Tajkhorshid⁶, Robert W. Evans⁷, Andrew R. Gorringer⁸, Anne B. Mason⁴, Alasdair C. Steven² & Susan K. Buchanan¹

***Neisseria* are obligate human pathogens causing bacterial meningitis, septicaemia and gonorrhoea. *Neisseria* require iron for survival and can extract it directly from human transferrin for transport across the outer membrane. The transport system consists of TbpA, an integral outer membrane protein, and TbpB, a co-receptor attached to the cell surface; both proteins are potentially important vaccine and therapeutic targets. Two key questions driving *Neisseria* research are how human transferrin is specifically targeted, and how the bacteria liberate iron from transferrin at neutral pH. To address these questions, we solved crystal structures of the TbpA–transferrin complex and of the corresponding co-receptor TbpB. We characterized the TbpB–transferrin complex by small-angle X-ray scattering and the TbpA–TbpB–transferrin complex by electron microscopy. Our studies provide a rational basis for the specificity of TbpA for human transferrin, show how TbpA promotes iron release from transferrin, and elucidate how TbpB facilitates this process.**

Neisseria comprise a large family of Gram-negative bacteria that colonize humans. Two family members, *Neisseria gonorrhoeae* and *Neisseria meningitidis*, are pathogens that invade the urogenital tract and nasopharynx, respectively, causing gonorrhoea, meningitis and other systemic infections. Although vaccines exist for bacterial meningitis, they have significant limitations¹ and are ineffective against serogroup B *N. meningitidis*. Currently, there are no vaccines to protect against gonococcal infections. The recent emergence of antibiotic-resistant strains² adds urgency to the need to develop more effective countermeasures for both pathogens.

Neisseria require iron for survival and virulence³. Unlike most Gram-negative bacteria, *Neisseria* do not make siderophores but instead extract iron directly from serum transferrin in the human host (TF). The neisserial transport system consists of two large surface proteins: TF-binding protein A (TbpA), a 100-kDa integral outer membrane protein belonging to the family of TonB-dependent transporters⁴; and TF-binding protein B (TbpB), an ~80-kDa co-receptor attached to the outer membrane by a lipid anchor (Supplementary Fig. 1). Both proteins are found in all clinical isolates of pathogenic *Neisseria*. TbpA binds apo and iron-containing transferrin with similar affinities, whereas TbpB only associates with iron-bound TF^{5,6}. Although TbpA can extract and import iron without TbpB, the process is considerably more efficient in the presence of the co-receptor^{7,8}. TbpA and TbpB induce bactericidal antibodies in mice against *N. meningitidis*^{9,10} and *N. gonorrhoeae*¹¹, making both proteins important vaccine targets. To elucidate how TbpA and TbpB function to bind human TF selectively and extract its tightly bound iron ($K_a = 10^{23} \text{ M}^{-1}$) at physiological pH, we combined an approach consisting of X-ray crystallography, small-angle X-ray scattering and electron microscopy to determine a model of the 260-kDa iron import complex from *N. meningitidis* strain K454 (serogroup B). Because *N. gonorrhoeae* strains FA1090 and FA19 express TbpA proteins that are 94% identical to the meningococcal protein, whereas the corresponding TbpB proteins

are 61% and 69% identical, respectively, our results are relevant to both pathogens.

Crystal structure of the TbpA–(apo)hTF complex

Structural characterization of the neisserial iron import machinery was initiated by crystallizing *N. meningitidis* TbpA with full-length, glycosylated apo human transferrin (hTF) and solving the structure to a resolution of 2.6 Å (Fig. 1, Supplementary Figs 2–4, 17 and Supplementary Table 1). Despite being significantly larger (~20%) than other structurally characterized TonB-dependent transporters⁴, TbpA retains the classic fold with a 22-strand transmembrane β -barrel encompassing a plug domain (Fig. 1a). Most of the additional mass is found in several extracellular loops which extend up to ~60 Å above the outer membrane. A plug loop implicated in iron uptake^{12,13} is unusually long and protrudes ~25 Å above the cell surface.

Human TF is a bilobal glycoprotein (~80 kDa) with a single ferric (Fe^{3+}) iron tightly bound within a cleft in each lobe (Fig. 1a and Supplementary Fig. 8). Each lobe of TF consists of two subdomains which form the cleft: N1, N2, C1 and C2. In the absence of iron, each lobe adopts an open conformation (Protein Data Bank (PDB) accession code 2HAV)¹⁴. To obtain the best model of neisserial iron import, we solved the structure of diferric hTF at 2.1 Å resolution (Fig. 2d and Supplementary Fig. 8). In our diferric structure, each lobe is found in a fully closed conformation, nearly identical to the diferric structures for both porcine (PDB code 1H76) and rabbit (PDB code 1JNF) TF.

When TbpA binds hTF, it sequesters ~2,500 Å² of buried surface, with 81 TbpA residues and 67 hTF residues participating in the interaction (Fig. 1a, c, e and Supplementary Table 2). TbpA binds exclusively to the C lobe of hTF, where electrostatic complementarity exists between the extracellular surface of TbpA (electropositive) and the C1 subdomain of hTF (electronegative) (Fig. 1b, d). Two notable features of the interface include: (1) the unusually long TbpA plug loop (residues 121–139) interacts directly with the C1 subdomain

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA. ²Laboratory of Structural Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA. ³Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. ⁴Department of Biochemistry, University of Vermont, College of Medicine, 89 Beaumont Avenue, Burlington, Vermont 05405, USA. ⁵Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. ⁶Department of Biochemistry and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. ⁷Metalloprotein Research Group, Division of Biosciences, School of Health Sciences and Social Care, Brunel University, Uxbridge, Middlesex UB8 3PH, UK. ⁸Health Protection Agency, Porton Down, Salisbury SP2 8NY, UK.

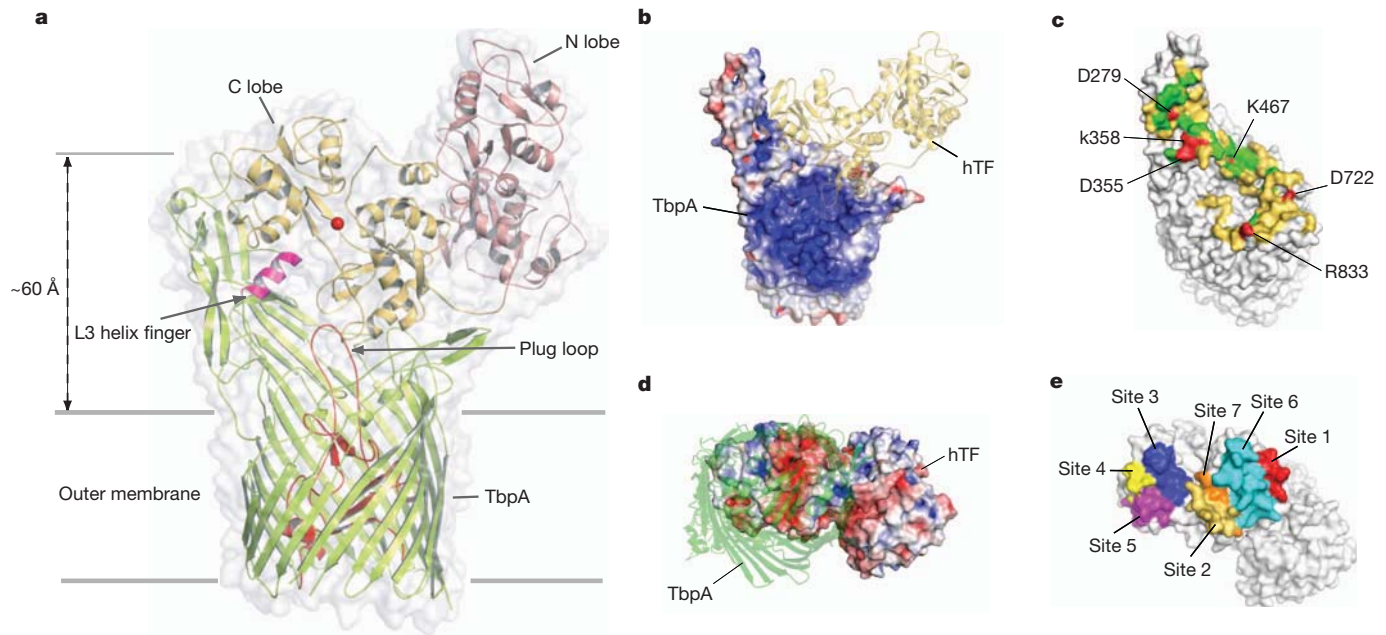


Figure 1 | Crystal structure of the TbpA–(apo)hTF complex. **a**, The TbpA β -barrel is lime, the plug is red, the helix finger is magenta. For hTF, the C lobe is gold and the N lobe is salmon. A ferric ion has been modelled into the C lobe as a red sphere. **b**, Electrostatic potential of TbpA viewed from the extracellular surface with hTF shown in gold ribbon. **c**, Residues of TbpA that bind hTF:

(Figs 1a, 2a, c and Supplementary Fig. 5), and (2) an α -helix in TbpA extracellular loop 3 (residues 351–361, the L3 ‘helix finger’) is inserted directly into the cleft of the C lobe between the C1 and C2 subdomains (Figs 1a, 2a, b and Supplementary Fig. 5). The interaction between TbpA and hTF was found to be relatively tolerant to point mutations in TbpA, as might be expected given the large binding interface (Supplementary Fig. 6).

The mechanism of species specificity of neisserial TbpA for hTF is unknown. In *in vitro* assays, gonococcal and meningococcal TbpA proteins have been shown to bind human TF, but not TF from cow, horse, rabbit, mouse, rat, sheep, duck or pig^{15–17}. In addition, mice

infected with *N. meningitidis* displayed a higher mortality rate when the iron source was Fe_2hTF rather than bovine TF¹⁸. From the TbpA–hTF crystal structure, seven sites spanning both the C1 and C2 subdomains of hTF participate in binding TbpA (Fig. 1e and Supplementary Fig. 7), with each site containing one or more residues unique to human TF.

Because TbpA shows limited sequence variation (Supplementary Fig. 5 and Supplementary Table 4), is present in all clinical isolates, and nearly all the interactions with hTF are mediated by extracellular loops of TbpA, we attempted to disrupt the TbpA–hTF interface to see if this would be a viable therapeutic strategy. Peptides from four loops

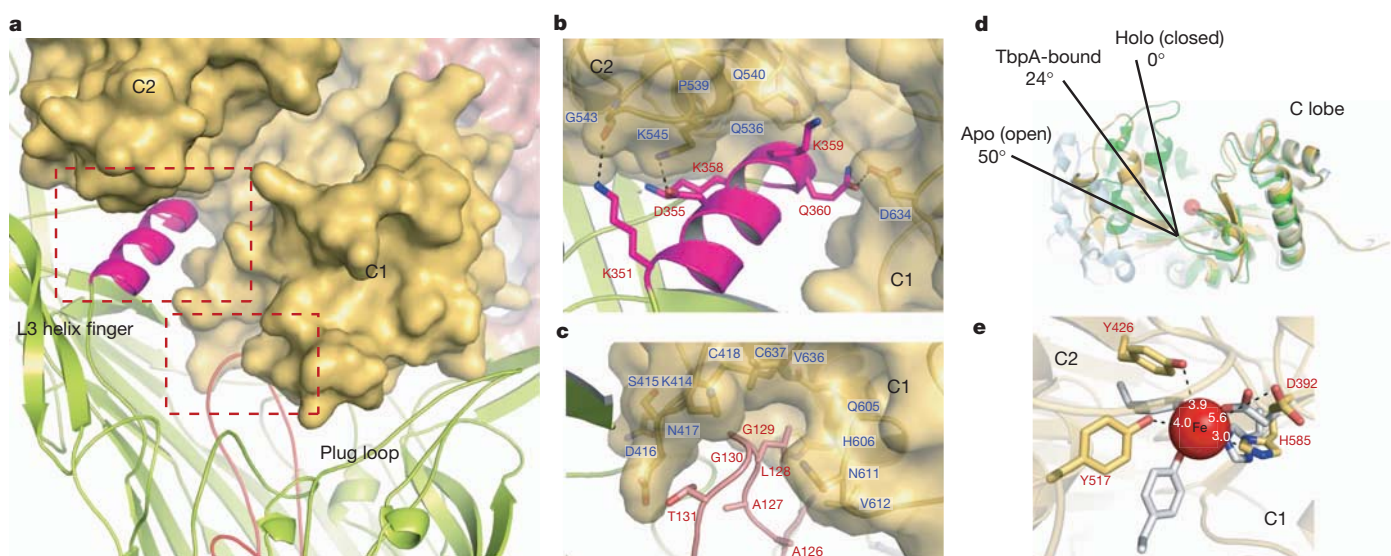


Figure 2 | TbpA distorts the iron coordination site in the hTF C lobe by inserting a helix from extracellular loop three. **a**, TbpA (green) inserts a helix finger from extracellular loop 3 (magenta) into the cleft of the hTF C lobe (gold). **b**, The helix finger interacts with the hTF C-lobe residues through main-chain and side-chain interactions. **c**, The long TbpA plug loop (pink) interacts

with residues from the C1 subdomain. **d**, Comparison of C-lobe conformations for holo (green), apo (grey, PDB code 2HAU), and TbpA-bound TF (gold). **e**, Superposition of residues coordinating iron in diferric hTF (grey) with the same residues in hTF when bound to TbpA (gold). Distances for the residues coordinating iron are shown for the TbpA-bound state.

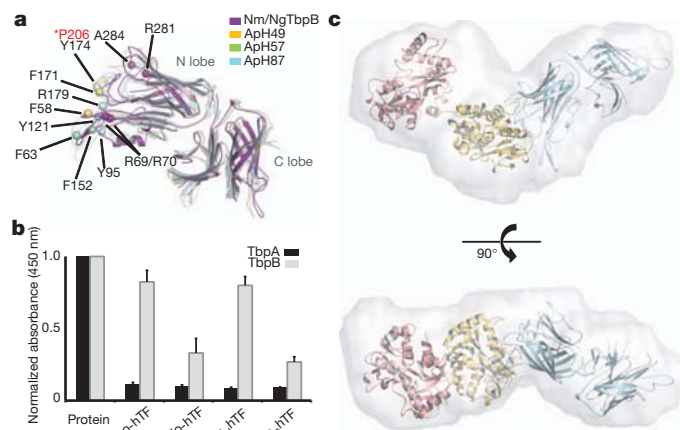


Figure 3 | SAXS analysis of the *N. meningitidis* TbpB-(holo)hTF complex. **a**, Superposition of TbpB from *N. meningitidis* with three TbpB structures from porcine pathogens. Whereas C lobes align closely, the N lobes show sequence and structural variability. Residues that diminish hTF binding when mutated are shown as spheres. Position 206 (proline in *N. meningitidis*) is critical for interaction with hTF. **b**, Competition ELISA showing relative binding affinities of TbpA and TbpB for apo-hTF, holo-hTF, hTF-Fe_N and hTF-Fe_C. Each experiment was performed in triplicate and data reported with standard errors. **c**, Fitting of the TbpB-(holo)hTF complex model into the averaged *ab initio* envelope was performed using Chimera. TbpB is shown in cyan, and hTF is shown in salmon (N lobe) and gold (C lobe).

(loops 3, 7, 11 and the plug loop) that make substantial contacts with hTF were used as antigens for polyclonal antibody development (Supplementary Table 3). All four antibodies reduced hTF binding (Supplementary Fig. 9). These results show that although the TbpA-hTF interface is extensive, reagents can be designed to disrupt it.

TbpA induces a conformational change in the hTF C lobe

In the full-length apo-hTF structure¹⁴, both the N and C lobes are in 'open' conformations, with 59.4° and 49.5° rotations required to align subdomains with diferric hTF (Fig. 2d and Supplementary Fig. 8). In the TbpA-hTF complex crystal structure, the N lobe is in the fully

open conformation. Notably, interaction with TbpA causes the C lobe to adopt a conformation midway between open and closed, with a 24° rotation required to align C1 and C2 subdomains with diferric hTF (Fig. 2d and Supplementary Fig. 10). The TbpA L3 helix finger is inserted into the cleft, where it interacts with D634 from the C1 subdomain and several residues from the C2 subdomain (Fig. 2a, b). The long TbpA plug loop also interacts with the surface of the C1 subdomain of hTF (Fig. 2c). These interactions induce a partial opening of the cleft in the hTF C lobe, thereby destabilizing the iron coordination site to facilitate the release of iron from the C lobe to TbpA. Figure 2e shows the residues coordinating iron in the hTF C-lobe structure and the significant increase in these distances when hTF binds TbpA. Such increases are clearly incompatible with tight binding of iron in the C lobe.

X-ray and SAXS structures of TbpB and TbpB-(holo)hTF

Although only TbpA can acquire iron from hTF, the reaction is enhanced by expression of the co-receptor TbpB^{7,8}, which preferentially binds holo-hTF⁶. To understand how TbpB facilitates iron extraction and uptake, we solved the structure of *N. meningitidis* TbpB (Fig. 3a, Supplementary Figs 12, 18 and Supplementary Table 1). TbpB consists of N and C lobes that are structurally similar, sharing an eight-strand β -barrel subdomain flanked by a four-strand 'handle' domain.

TbpB proteins from different isolates vary substantially in size and sequence (Supplementary Fig. 11), but the overall fold is conserved. Our structure aligns closely with three TbpB structures from porcine pathogens^{19,20} and shows that sequence and conformational variations are found primarily in the N lobe (Fig. 3a and Supplementary Fig. 12b). Specifically, residues affecting TF binding lie on the distal surface of the N lobe^{19–22} and confer much of the TF species specificity (Fig. 3a). We made point mutants at four sites on this surface that reduced or abolished binding to hTF (Fig. 3a and Supplementary Fig. 12). An analysis of our mutants, and those reported previously, demonstrates that the major site of interaction lies in the N lobe.

To clarify the interactions of purified TbpA and TbpB with hTF, an enzyme-linked immunosorbent assay (ELISA) was used to probe binding to Fe₂hTF, monoferric hTF with iron only in the N lobe (Fe_N), or in the C lobe (Fe_C), and apo-hTF (incapable of binding iron

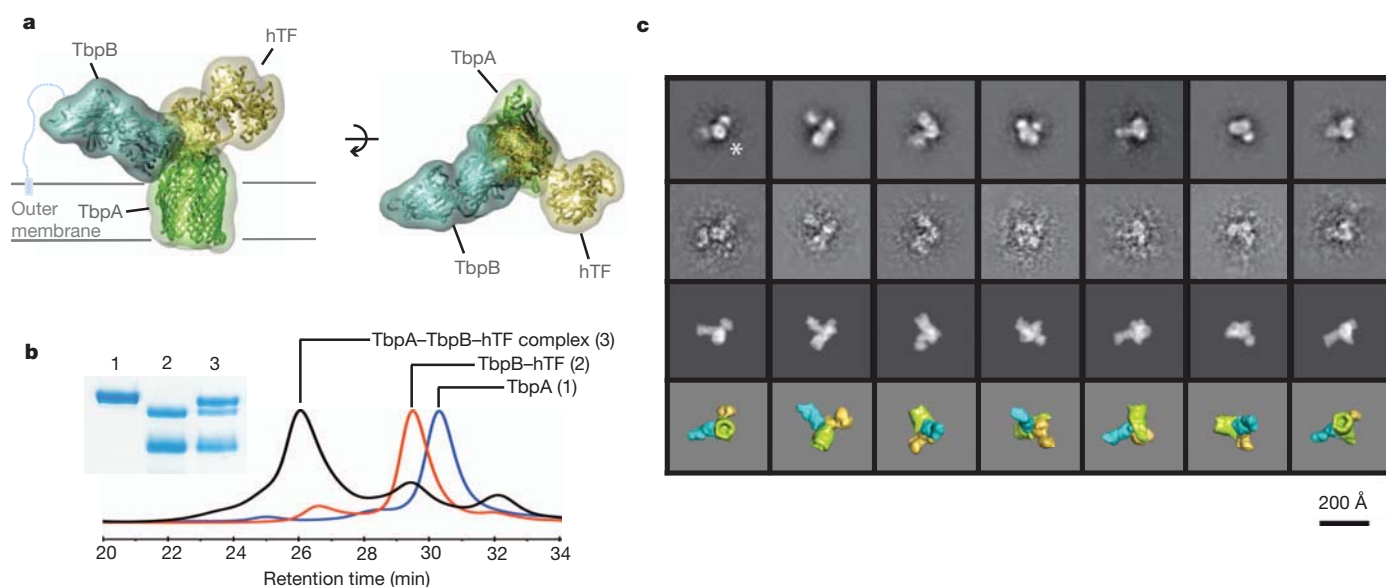


Figure 4 | Analysis of the TbpA-TbpB-(holo)hTF triple complex by negative stain electron microscopy. **a**, Model of the triple complex. TbpA is green, TbpB is light blue and hTF is gold. **b**, Purification of the triple complex by size-exclusion chromatography. The Coomassie-stained SDS gel shows purified TbpA in lane 1, the TbpB-(holo)hTF complex in lane 2, and the TbpA-

TbpB-(holo)hTF complex in lane 3. **c**, Row 1, set of seven non-redundant class averages of negatively stained complexes; row 2, examples of individual images assigned to the respective classes; row 3, re-projections of the model in the corresponding orientations, band-limited to 15 Å resolution; row 4, surface renderings of the band-limited model in the corresponding orientations.

in either lobe)²³. Consistent with earlier studies using apo- or holo-hTF⁶, TbpA binds all four hTFs with equal affinity regardless of the iron status of either lobe (Fig. 3b). In contrast, TbpB has a strong preference for hTF constructs with iron bound in the C lobe, regardless of the coordination state of the N lobe. These experiments clearly show that, at least *in vitro*, hTF interacts with TbpA and TbpB solely through the C lobe and is not affected by the presence or absence of iron in the N lobe. Our results indicate that *Neisseria* cannot use the entire serum TF iron supply and that the primary function of TbpB is to select and concentrate on the cell surface only those forms of TF that are able to provide iron to the bacterium.

Because TbpB primarily binds the C lobe of hTF through its N lobe, we performed steered molecular docking for the TbpB–hTF complex based on previous docking studies for the porcine complex²⁰ and on our mutagenesis results. We collected small-angle X-ray scattering (SAXS) data on the TbpB–(holo)hTF complex (Supplementary Figs 12 and 13) and used GASBOR²⁴ to construct the SAXS envelope (Fig. 3c and Supplementary Fig. 13). The resulting molecular envelope describes the spatial arrangement of TbpB and hTF, and was used to fit the TbpB–hTF complex structure. Binding TbpB to hTF buries ~1,300 Å² of surface area and, notably, uses a region of the hTF C lobe distinct from the site where TbpA binds.

Structure of the triple complex by single-particle EM

On the basis of the TbpA–(apo)hTF crystal structure and the SAXS solution structure of the TbpB–(holo)hTF complex, we formed an *in silico* model for the TbpA–TbpB–(holo)hTF triple complex by

superposing the two complexes along the C1 subdomain of hTF (Fig. 4a). To test this model, we assembled the triple complex from its components (Fig. 4b) and visualized the resulting particles by negative staining electron microscopy (EM) (Fig. 4c and Supplementary Fig. 14). A set of 4,240 particles was subjected to a reference-free classification to identify subsets of like images, representing molecules viewed in the same orientation; the members of each class were then averaged to suppress noise. Several of the class averages show a central density, ~45 Å across, to which two small globular densities are appended at points about 120° apart around its periphery (for example, Fig. 4c (asterisk) and Supplementary Fig. 14). A plausible interpretation is that the central density corresponds to the β-barrel domain of TbpA and the two appended densities to TbpB and hTF (Supplementary Fig. 14), in agreement with our model for the triple complex.

Iron extraction and import

The X-ray, SAXS and EM structures support a consistent arrangement for the TbpA–TbpB–(holo)hTF complex. Although TbpA and TbpB each bind hTF tightly through the C lobe, they have unique, non-overlapping binding sites (Fig. 5a). A consequence of the assembly of the triple complex is the formation of an enclosed chamber (volume of ~1,000 Å³) at the union of the three protein components, which sits directly above the plug domain of TbpA (Fig. 5b). This chamber may serve two important roles for iron acquisition by the bacteria: (1) prevent diffusion of iron released from hTF; and (2) guide the iron towards the β-barrel domain of TbpA for subsequent transport.

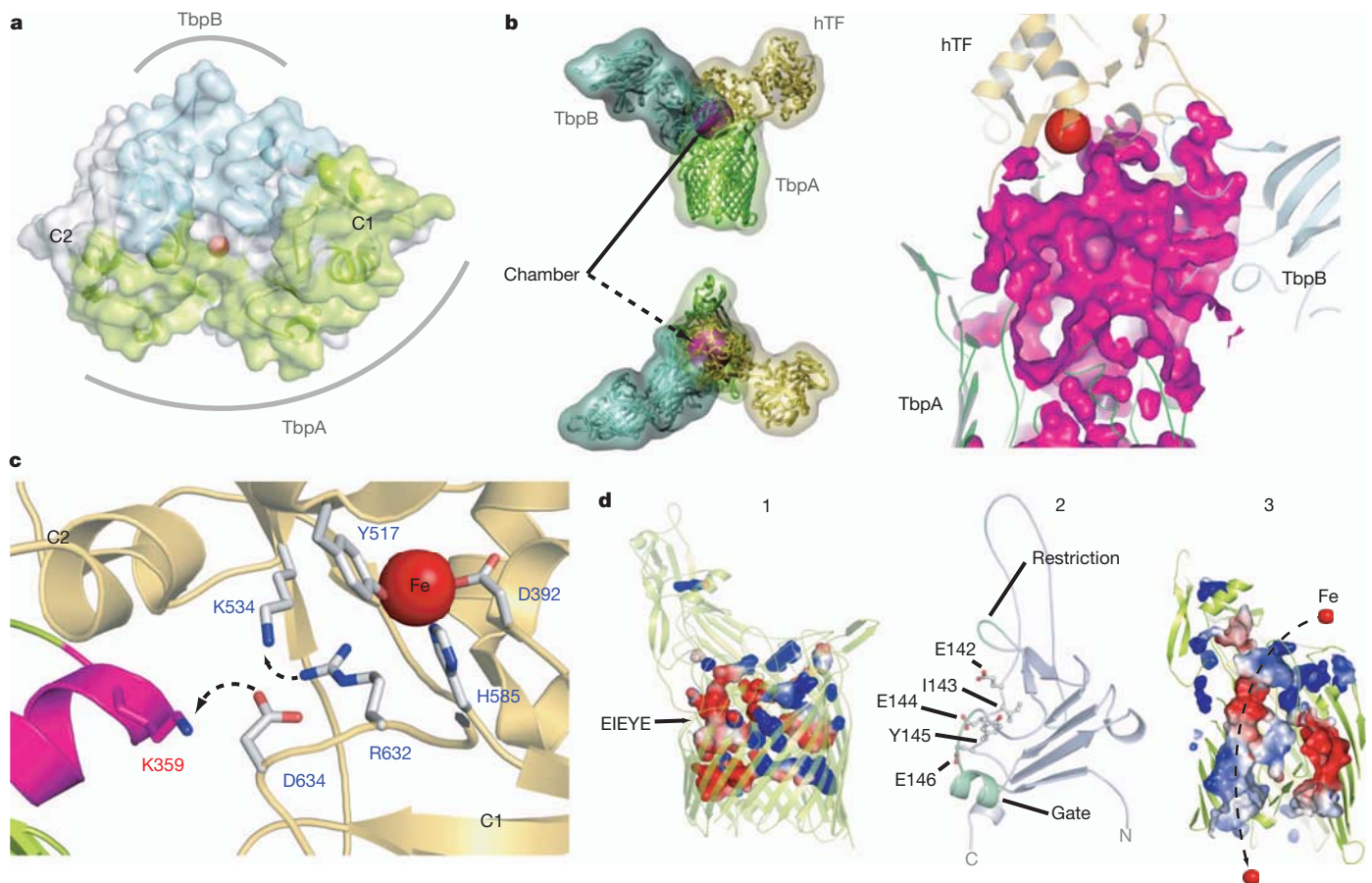


Figure 5 | Mechanism for iron import. **a**, Binding surfaces of TbpA (green) and TbpB (cyan) mapped onto the hTF C lobe. **b**, Enclosed chamber formed by TbpA–TbpB–(holo)hTF (left, magenta sphere). A cutaway view (right) from inside the chamber illustrates the proximity of the iron (red). **c**, Model for iron release. Conserved K359 in the L3 helix finger is positioned to interact with

residues that regulate iron release in eukaryotic iron uptake. **d**, Import of iron through TbpA. 1, an electrostatic surface depicts cavities between the TbpA barrel and plug domain; 2, plug domain constrictions close the tunnel; 3, molecular dynamics simulations show removal of constrictions upon interaction with TonB.

A plausible mechanism for iron extraction from hTF is shown in Fig. 5c. Insertion of the TbpA L3 helix finger into the cleft between the C1 and C2 subdomains of hTF positions a conserved lysine (TbpA K359; Supplementary Fig. 4) near the hTF triad of charged residues (hTF K534, R632 and D634) that has been implicated in iron release from the C lobe²⁵. TbpA K359 is perfectly situated to interact with D634, which would disrupt the charge neutralization it normally provides to the two basic triad residues K534 and R632. This potential charge repulsion between the hTF C1 and C2 subdomains could induce cleft opening (as observed in the TbpA-(apo)hTF crystal structures), resulting in distortion of the C-lobe iron-binding site and subsequent iron release. Notably, a recent study indicates that an hTF D634A mutant has a rate of iron release that is 80-fold faster at pH 5.6 than the control under the same conditions²⁶.

To investigate iron transport across the outer membrane, steered molecular dynamics was used to simulate interactions between TbpA and TonB (Supplementary Fig. 15). In the ground state structure, a large, highly negative transmembrane cavity is located between the barrel wall and the plug domain, but access is restricted on the extracellular side by residues 91–96 (restriction loop) and on the periplasmic side by residues 65–71 (helical gate) from the plug domain (Fig. 5d). When force (designed to mimic interaction with TonB) is applied to the plug domain, it sequentially unfolds beginning with removal of the helical gate followed by the restriction loop, producing an unobstructed pathway from the extracellular space to the periplasm (Supplementary Movie 1). This pathway is lined by the EIEYE motif of the plug domain¹³, which contains multiple oxygen donor groups that could transiently bind iron as it is transported through TbpA.

Concluding remarks

Humans and bacteria have each developed unique strategies to acquire iron from serum transferrin^{27,28} (Supplementary Fig. 16). Our TbpA–TbpB–hTF X-ray, SAXS and EM structures indicate a mechanism for bacterial uptake of iron with the following characteristics: (1) a large TbpA–hTF binding interface with many human-TF-specific interactions; (2) iron removal from the hTF C lobe by insertion of a helical element from TbpA into the iron-binding cleft; and (3) iron transport across the outer membrane after TonB-dependent conformational changes in the TbpA plug domain. This system allows efficient extraction of iron despite the extremely high affinity of hTF bound iron at neutral pH. The TbpB co-receptor, which is tethered to the cell surface by a long, unstructured polypeptide chain, is able to attract and preferentially bind hTF with iron in the C lobe, thereby increasing the efficiency of the system. Crucial to the mechanism, TbpA and TbpB associate with different regions of the hTF C lobe, creating an enclosed chamber above the plug domain to ensure that iron is efficiently sequestered and directionally transported through the TbpA barrel (Supplementary Movie 2). Finally, as TbpA and TbpB are surface-exposed, antigenic and required for neisserial infections²⁹, our structures provide the necessary information for structure-based vaccine and drug design³⁰.

METHODS SUMMARY

The TbpA–(apo)hTF complex was crystallized from TbpA expressed in *Escherichia coli* and apo-hTF purchased from Sigma-Aldrich. For the TbpA–(apo)hTF C-lobe structure, hTF incorporating a protease cleavage site between N and C lobes was expressed in BHK cells and purified as described³¹. Full-length N-His-tagged hTFs including holo, authentic apo, and both monoferric forms were expressed in BHK cells and purified as described²³. The diferric hTF structure was solved using protein purchased from Sigma-Aldrich. The hTF C-lobe structure was solved using hTF C lobe from a construct containing a TEV protease site between the N and C lobes, expressed in BHK cells, and purified³². TbpB was expressed in *E. coli*. X-ray data were collected at GM/CA and SER-CAT beamlines of the Advanced Photon Source synchrotron. SAXS data was collected on beamline BL4-2 at the Stanford Synchrotron Radiation Lightsource. EM data were collected on a CM120–LaB6 electron microscope (FEI), operating at 120 kV. Molecular dynamics simulations were performed using the program NAMD³³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 September 2011; accepted 9 January 2012.

Published online 12 February 2012.

- Deasy, A. & Read, R. C. Challenges for development of meningococcal vaccines in infants and children. *Expert Rev. Vaccines* **10**, 335–343 (2011).
- Centers for Disease Control and Prevention. Cephalosporin susceptibility among *Neisseria gonorrhoeae* isolates—United States, 2000–2010. *Morbid. Mortal. Weekly Rep.* **60**, 873–877 (2011).
- Grifantini, R. et al. Identification of iron-activated and -repressed Fur-dependent genes by transcriptome analysis of *Neisseria meningitidis* group B. *Proc. Natl Acad. Sci. USA* **100**, 9542–9547 (2003).
- Noiraj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).
- Boulton, I. C. et al. Transferrin-binding protein B isolated from *Neisseria meningitidis* discriminates between apo and diferric human transferrin. *Biochem. J.* **334**, 269–273 (1998).
- Krell, T. et al. Insight into the structure and function of the transferrin receptor from *Neisseria meningitidis* using microcalorimetric techniques. *J. Biol. Chem.* **278**, 14712–14722 (2003).
- Anderson, J. E., Sparling, P. F. & Cornelissen, C. N. Gonococcal transferrin-binding protein 2 facilitates but is not essential for transferrin utilization. *J. Bacteriol.* **176**, 3162–3170 (1994).
- Irwin, S. W., Averil, N., Cheng, C. Y. & Schryvers, A. B. Preparation and analysis of isogenic mutants in the transferrin receptor protein genes, *tbpA* and *tbpB*, from *Neisseria meningitidis*. *Mol. Microbiol.* **8**, 1125–1133 (1993).
- Rokbi, B. et al. Evaluation of recombinant transferrin-binding protein B variants from *Neisseria meningitidis* for their ability to induce cross-reactive and bactericidal antibodies against a genetically diverse collection of serogroup B strains. *Infect. Immun.* **65**, 55–63 (1997).
- Weynants, V. E. et al. Additive and synergistic bactericidal activity of antibodies directed against minor outer membrane proteins of *Neisseria meningitidis*. *Infect. Immun.* **75**, 5434–5442 (2007).
- Price, G. A., Masri, H. P., Hollander, A. M., Russell, M. W. & Cornelissen, C. N. Gonococcal transferrin binding protein chimeras induce bactericidal and growth inhibitory antibodies in mice. *Vaccine* **25**, 7247–7260 (2007).
- Yost-Daljev, M. K. & Cornelissen, C. N. Determination of surface-exposed, functional domains of gonococcal transferrin-binding protein A. *Infect. Immun.* **72**, 1775–1785 (2004).
- Noto, J. M. & Cornelissen, C. N. Identification of TbpA residues required for transferrin-iron utilization by *Neisseria gonorrhoeae*. *Infect. Immun.* **76**, 1960–1969 (2008).
- Wally, J. et al. The crystal structure of iron-free human serum transferrin provides insight into inter-lobe communication and receptor binding. *J. Biol. Chem.* **281**, 24934–24944 (2006).
- Cornelissen, C. N., Biswas, G. D. & Sparling, P. F. Expression of gonococcal transferrin-binding protein 1 causes *Escherichia coli* to bind human transferrin. *J. Bacteriol.* **175**, 2448–2450 (1993).
- Schryvers, A. B. & Morris, L. J. Identification and characterization of the transferrin receptor from *Neisseria meningitidis*. *Mol. Microbiol.* **2**, 281–288 (1988).
- Stokes, R. H., Oakhill, J. S., Joannou, C. L., Gorrings, A. R. & Evans, R. W. Meningococcal transferrin-binding proteins A and B show cooperation in their binding kinetics for human transferrin. *Infect. Immun.* **73**, 944–952 (2005).
- Schryvers, A. B. & Gonzalez, G. C. Comparison of the abilities of different protein sources of iron to enhance *Neisseria meningitidis* infection in mice. *Infect. Immun.* **57**, 2425–2429 (1989).
- Calmettes, C. et al. Structural variations within the transferrin binding site on transferrin-binding protein B, TbpB. *J. Biol. Chem.* **286**, 12683–12692 (2011).
- Moraes, T. F., Yu, R. H., Strynadka, N. C. & Schryvers, A. B. Insights into the bacterial transferrin receptor: the structure of transferrin-binding protein B from *Actinobacillus pleuropneumoniae*. *Mol. Cell* **35**, 523–533 (2009).
- Cornelissen, C. N., Anderson, J. E. & Sparling, P. F. Characterization of the diversity and the transferrin-binding domain of gonococcal transferrin-binding protein 2. *Infect. Immun.* **65**, 822–828 (1997).
- Silva, L. P. et al. Conserved interaction between transferrin and transferrin-binding proteins from porcine pathogens. *J. Biol. Chem.* **286**, 21353–21360 (2011).
- Mason, A. B. et al. Expression, purification, and characterization of authentic monoferric and apo-human serum transferrins. *Protein Expr. Purif.* **36**, 318–326 (2004).
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
- Halbrooks, P. J. et al. Investigation of the mechanism of iron release from the C-lobe of human serum transferrin: mutational analysis of the role of a pH sensitive triad. *Biochemistry* **42**, 3701–3707 (2003).
- Steere, A. N., Byrne, S. L., Chasteen, N. D. & Mason, A. B. Kinetics of iron release from transferrin bound to the transferrin receptor at endosomal pH. *Biochim. Biophys. Acta* <http://dx.doi.org/10.1016/j.bbagen.2011.06.003> (2011).
- Cheng, Y., Zak, O., Aisen, P., Harrison, S. C. & Walz, T. Structure of the human transferrin receptor-transferrin complex. *Cell* **116**, 565–576 (2004).
- Eckenroth, B. E., Steere, A. N., Chasteen, N. D., Everse, S. J. & Mason, A. B. How the binding of human transferrin primes the transferrin receptor potentiating iron release at endosomal pH. *Proc. Natl Acad. Sci. USA* **108**, 13089–13094 (2011).

29. Hobbs, M. M. *et al.* Experimental gonococcal infection in male volunteers: cumulative experience with *Neisseria gonorrhoeae* strains FA1090 and MS11mkC. *Front. Microbiol.* **2**, 123 (2011).
30. Scarselli, M. *et al.* Rational design of a meningococcal antigen inducing broad protective immunity. *Sci. Transl. Med.* **3**, 91ra62 (2011).
31. Zak, O. & Aisen, P. A new method for obtaining human transferrin C-lobe in the native conformation: preparation and properties. *Biochemistry* **41**, 1647–1653 (2002).
32. Steere, A. N. *et al.* Properties of a homogeneous C-lobe prepared by introduction of a TEV cleavage site between the lobes of human transferrin. *Protein Expr. Purif.* **72**, 32–41 (2010).
33. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements N.N., N.C.E., M.O., E.B. and S.K.B. are supported by the Intramural Research Program of the NIH, National Institute of Diabetes and Digestive and Kidney Diseases. M.O. was initially funded by an EPSRC Research Committee Studentship awarded to S.K.B. and R.W.E. N.M. and A.C.S. are supported by the Intramural Research Program of the NIH, National Institute of Arthritis and Musculoskeletal and Skin Diseases. A.B.M. was supported in part by USPHS grant R01-DK21739. A.N.S. is funded by an AHA Predoctoral Fellowship (10PRE4200010). E.T. acknowledges NIH support by R01-GM086749, U54-GM087519 and P41-RR05969. All the simulations were performed using TeraGrid resources (MCA06N060). We thank the respective staffs at the Southeast Regional Collaborative Access Team (SER-CAT) and General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) beamlines at the Advanced Photon Source, Argonne National Laboratory for their assistance during data

collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38 (SER-CAT), and by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357 (GM/CA-CAT). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program.

Author Contributions N.N., N.C.E., M.O. and S.K.B. expressed, purified and crystallized TbpA, TbpB and various hTFs. N.N. solved all crystal structures and the SAXS structure and analysed all data. A.B.M. and A.N.S. designed and purified apo-hTF, holo-hTF, hTF-Fe_N and hTF-Fe_C for binding experiments with TbpA and TbpB; they also expressed and purified hTF C lobe for the corresponding structure (PDB code 3SKP). P.A. and O.Z. expressed and purified hTF C lobe for the TbpA-(apo)hTF C-lobe structure (PDB code 3V89). N.M. and A.C.S. designed, conducted and analysed EM experiments. E.T. and J.G. designed, conducted and analysed molecular dynamics simulations. E.B. participated in the data collection and analysis of the SAXS data. R.W.E., A.R.G. and S.K.B. conceived and designed the original project. N.N. and S.K.B. wrote the manuscript.

Author Information Coordinates and structure factors for TbpA-(apo)hTF, TbpA-(apo)hTF C lobe, diferric hTF, apo-hTF C lobe and TbpB are deposited in the Protein Data Bank under accession codes 3V8X, 3V89, 3V83, 3SKP and 3V8U, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K.B. (skbuchan@helix.nih.gov).

METHODS

Cloning, expression and purification of TbpA. The *N. meningitidis* TbpA sequence from strain K454 (B15:P1.7,16) was subcloned into pET20b (Novagen) containing an N-terminal 10×-His tag. TbpA mutants were created using site-directed mutagenesis using QuikChange (Stratagene). For structural studies, mutation of M889 to Tyr improved expression levels. TbpA was expressed in BL21(DE3) cells at 20 °C without induction in terrific broth (TB) and carbenicillin. Expression for the mutants followed the same protocol.

For purification, cells were re-suspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 1 mM MgCl₂, 10 µg ml⁻¹ DNaseI, 100 µg ml⁻¹ AEBSF) and lysed by two passages through an Emulsiflex C3 (Avestin) homogenizer at 4 °C. The lysate was centrifuged at 12,000g for 10 min to remove unlysed cells and the supernatant was incubated with 2% Triton X-100 for 30 min at room temperature. The mixture was centrifuged at 160,000g for 90 min at 4 °C. The membrane pellets were re-suspended in 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 20 mM imidazole and solubilized by constant stirring using 5% Elugent for 16 h at 4 °C. Solubilized membranes were centrifuged at 265,000g for 60 min at 4 °C and the supernatant filtered and applied to a 15-ml Ni-NTA column. TbpA was eluted using 250 mM imidazole. Peak fractions were concentrated and applied to an S-300HR Sephacryl size exclusion column (GE Healthcare) using 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.8% C₈E₄ and 0.02% NaN₃. Peak fractions were verified using SDS-PAGE and western blot analysis using an anti-His monoclonal antibody (Sigma).

Cloning, expression and purification of TbpB. The TbpB sequence (starting at residue L22) from *N. meningitidis* K454 was codon optimized and synthesized by GenScript and subsequently subcloned into a pET28b vector (Novagen). TbpB was expressed in T7-Express cells (NEB) at 37 °C with IPTG induction at an optical density at 600 nm of 0.75–1.0 with continued expression for 4 h. Mutants were expressed using the same protocol.

For purification, cells were harvested and re-suspended in 5 ml PBS per gram of cell paste and supplemented with 10 µg ml⁻¹ AEBSF and 100 µg ml⁻¹ DNaseI. Cells were lysed by French press and then centrifuged for 45 min at 38,400g. The supernatant was applied to a Ni-NTA column and washed with 10 column volumes of PBS. A final wash was performed with PBS containing 20 mM imidazole before elution with PBS/250 mM imidazole. Eluted protein was then dialysed against PBS overnight at 4 °C. For constructs where the His tag was removed, TEV-HIS protease was added, the sample was dialysed and then passed through a second Ni-NTA column to remove uncleaved protein and protease. Finally, samples were purified by size-exclusion chromatography in PBS/0.02% NaN₃.

Crystallization and data collection. For crystallization of the TbpA-hTF complex, apo-human transferrin (Sigma) was mixed with TbpA at a 2:1 ratio and incubated on ice for 1 h. The complex was isolated using Sephacryl S300HR chromatography equilibrated with 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10 mM Na-citrate, 1 mM EDTA, 0.8% C₈E₄ (Anatrace) and 0.02% NaN₃. Fractions corresponding to the TbpA-hTF complex were verified using SDS-PAGE, pooled and concentrated to 10 mg ml⁻¹. Heptane-1,2,3-triol was added to 3% final concentration, incubated on ice for 30 min and then the protein sample was filtered before crystallization. Sparse matrix screening was performed using a TTP Labtech Mosquito crystallization robot using hanging-drop vapour diffusion and plates incubated at 21 °C. The best crystals were grown in 24-well Linbro plates (Hampton Research) from 20% Peg3350 and 200 mM BaBr₂. Data were collected at the SER-CAT beamline of the Advanced Photon Source of Argonne National Laboratory and data processed using HKL2000³⁴. The space group was *P*2₁2₁2₁ with one mol per asymmetric unit (ASU) and final cell parameters *a* = 91.014, *b* = 129.362, *c* = 198.589, α = 90.00, β = 90.00, γ = 90.00.

For crystallization of the TbpA-hTF C-lobe complex, hTF C lobe³¹ was mixed in a 2:1 ratio with TbpA and the complex isolated by Sephacryl S300HR chromatography using 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 0.1% LDAO and 0.02% NaN₃. Final crystal conditions consisted of 21% PEG 1000, 100 mM sodium acetate buffer (pH 4.8), 200 mM NaCl, 0.1% LDAO and 3% heptane-1,2,3-triol. Data were collected and processed as described for the TbpA-hTF complex. The space group was *P*2₁ with one mol per ASU with final cell parameters *a* = 58.055, *b* = 107.592, *c* = 130.721, α = 90.00, β = 94.48, γ = 90.00.

TbpB was crystallized from a 10 mg ml⁻¹ solution with 2.0 M NaCl and 2.0 M ammonium sulphate. Data were collected at the GM/CA CAT beamline of the Advanced Photon Source of Argonne National Laboratory and data were processed using HKL2000³⁴. The space group was *P*2₁ with two molecules per ASU with final cell parameters *a* = 75.288, *b* = 82.761, *c* = 111.882, α = 90.00, β = 105.95, γ = 90.00.

For diferric hTF crystallization, 100 mg of holo-hTF (Sigma) was solubilized and further purified by Sephacryl S300HR chromatography using 20 mM Tris-HCl, pH 7.5 and 200 mM NaCl. The protein was then concentrated to ~50 mg ml⁻¹

and crystallized using 100 mM HEPES, pH 7.5, 1.6 M ammonium sulphate, and 2% PEG 1000, with red-tinted crystals appearing only after several months and being extremely sensitive to even slight temperature changes. Drops containing the crystals were quickly hydrated with 3.4 M ammonium sulphate immediately before being flash cooled in liquid nitrogen and stored for data collection. Data were collected at the SER-CAT beamline of the Advanced Photon Source of Argonne National Laboratory and data were processed using HKL2000³⁴. The space group is *C*2 with six molecules per ASU and final cell parameters *a* = 254.53, *b* = 173.00, *c* = 150.15, α = 90.00, β = 123.26, γ = 90.00.

For hTF C-lobe crystallization, (holo)C lobe³² was mixed with excess TbpB N lobe and the complex isolated by size-exclusion chromatography as above in 25 mM Tris pH 8.0, 200 mM NaCl. The complex was concentrated to ~10 mg ml⁻¹ and broad screening performed using a Mosquito crystallization robot. Several crystallization conditions were observed; however, none was red in colour as might be expected for iron-bound crystals and most contained citrate, which is a known iron chelator. Data were collected and analysed as for TbpB. The space group was *I*422 with 1 mol of hTF C lobe per ASU with final cell parameters *a* = 95.847, *b* = 95.847, *c* = 204.140, α = 90.00, β = 90.00, γ = 90.00. No TbpB N lobe was present in the crystals.

Structure determination. For TbpA-hTF, we were unable to collect useful heavy atom derivatives for experimental phasing, and selenomethionine-substituted TbpA protein yields were not sufficient for crystallization. We eventually used molecular replacement in PHASER-CCP4³⁵ to solve the TbpA-hTF complex structure. Here, we first searched for each of the two domains (N lobe and C lobe) of hTF using the deposited coordinates (PDB code 2HAV), which produced good solutions with *Z*-scores above 8. However, although the electron density for the hTF molecule was reasonable, the electron density for TbpA was poor and could not be used for model building. Our attempts at using known TonB-dependent transporter structures as search models (barrel and plug, together and individually) were unsuccessful (low *Z*-scores and LLG scores). We then aligned the TbpA sequence to our structure-based sequence alignment reported in our recent review⁴ and found that TbpA contained many conserved regions characteristic of TonB-dependent transporters. Using the alignment between TbpA and its closest relative, FhuA (10% identity, ClustalW), and trimming the extracellular loops, 500 models within a root mean squared deviation of 5 Å were produced using Modeller (Accelrys). Each of these models was then used for molecular replacement within PHASER-CCP4³⁵, with two of them producing *Z*-scores above 8. The solution with the highest LLG (containing both hTF and the TbpA model) was refined in PHENIX³⁶ producing an initial *R*/*R*_{free} value of 0.43/0.48. Further model building was performed using COOT³⁷ and subsequent refinement done in PHENIX³⁶ and BUSTER-TNT³⁸. During the final states of refinement, extra density was observed which was mapped to residues N413 and N611 of hTF, both of which are reported as possible N-linked glycosylation sites. Therefore, N-linked glycans were built for these two residues. The final structure was solved to 2.60 Å with *R*/*R*_{free} values of 0.22/0.28. The TbpA-hTF C-lobe structure was solved by molecular replacement using the coordinates from the TbpA-hTF (full-length) structure reported here. Two search models were formed, one for TbpA and one for hTF C lobe. PHASER-CCP4³⁵ was used for molecular replacement and subsequent refinement performed using PHENIX³⁶. The structure was solved to 3.1 Å resolution with final *R*/*R*_{free} values of 0.24/0.29.

The TbpB structure was solved by molecular replacement using PDB code 3HOL. An initial model was created using the Swiss Model server³⁹ that was subsequently divided into four different search domains. PHASER-CCP4³⁵ was used for molecular replacement and subsequent refinement performed using PHENIX³⁶. The structure was solved to 2.40 Å resolution with final *R*/*R*_{free} values of 0.25/0.30.

The diferric hTF crystal structure was solved by molecular replacement using Phaser-CCP4³⁵. Search models for the N lobe and C lobe were created separately with the program Chainsaw (CCP4) using the existing diferric porcine TF coordinates (PDB code 1H76). Six copies of each lobe (six molecules of hTF total) were found in the ASU and the iron sites were easily observed in the difference density. These iron sites were further verified in an anomalous difference electron density map. Refinement was performed using PHENIX³⁶ and the structure was solved to 2.1 Å resolution with final *R*/*R*_{free} values of 0.19/0.23.

The non-glycosylated hTF-C-lobe structure was solved by molecular replacement using PDB code 2HAU. An initial search model was formed by truncating the N-lobe domain. PHASER-CCP4³⁵ was used for molecular replacement and subsequent refinement performed using PHENIX³⁶. The structure was solved to 1.7 Å resolution with final *R*/*R*_{free} values of 0.17/0.19. For all structures, figures were made with PyMOL (Schrodinger) or Chimera⁴⁰ and annotated and finalized with Adobe Illustrator.

Dot blots. Whole cells ($2\ \mu\text{l}$, $0.01\ \text{g ml}^{-1}$) and cell lysates (unmodified for TbpB samples, or incubated for 3 h with 2 mM EDTA and 1% DDM at room temperature for TbpA samples) were spotted onto nitrocellulose membrane and allowed to dry at room temperature. The membranes were then blocked with PBST 2% BSA for 15 min, washed and probed with HRP-conjugated hTF (1:1,000) (Jackson ImmunoResearch) for 15 min. The membrane was then washed and imaged using the colorimetric substrate 3,3'-diaminobenzidine (Sigma) where the appearance of a red dot indicated specific binding of the hTF–HRP conjugate. The results from the mutants were compared to wild-type Tbp to determine their effect on hTF binding.

ELISA. Whole cells ($100\ \mu\text{l}$ at $10\ \text{mg ml}^{-1}$ or $1\ \text{mg ml}^{-1}$ in PBS) of wild-type TbpA, empty vector control (pET20b) and TbpA mutants were added to a NUNC polystyrene 96-well plate (Fisher Scientific) and incubated at 37°C overnight. Wells were washed $2\times$ with PBST and then blocked with PBST 2% BSA for 30 min and probed with hTF–HRP (1:1,000) for 15 min. Wells were washed $2\times$ in PBST, $2\times$ in PBS, and then developed using $100\ \mu\text{l}$ 3,3',5,5'-tetramethylbenzidine substrate (TMB, Sigma) for 5 min and terminated using Stop solution (Sigma). Absorbances of each well were determined using a BioRad iMark plate reader at 450 nm and data normalized and compared to wild-type TbpA. Each experiment was performed in triplicate and data reported with standard errors.

Antibody blocking assays. Using the TbpA–hTF crystal structure reported here, we designed four different peptides based on four loops from TbpA (loops 3, 7, 11 and plug loop) to be used as antigens for polyclonal antibody development (Precision Antibody). A fifth polyclonal antibody was developed using purified full-length TbpA ($1\times$ PBS 7.4, 0.1% DDM). An ELISA was designed to probe whether or not these antibodies could block hTF binding. Here, TbpA–His (20 ng) was incubated for 20 min in a final volume of $100\ \mu\text{l}$ either alone or in the presence of each antibody (1:20) individually in PBS containing 0.05% Cymal-6 (Anatrace). In addition, we tested the antibodies that targeted TbpA loops in combinations to determine if an additive effect could be observed. Each sample was then transferred to a 96-well Ni-NTA Agarose HiSorb plate (Qiagen) and incubated for 30 min and washed $2\times$ with PBST + 0.05% Cymal-6. Assays were performed as described in the previous section. In a second set of ELISAs, TbpA–His was first bound to the Ni-NTA Agarose HiSorb plate before incubation with antibodies. Results were analysed and initial graphs made using Microsoft Excel. The graphs were then imported, annotated and finalized with Adobe Illustrator.

Protease accessibility of TbpA and TbpA mutants. To confirm that TbpA and the TbpA mutants were being properly presented at the surface of the bacteria, we treated whole cells with trypsin ($5\ \mu\text{g ml}^{-1}$ final concentration) for 15 min at room temperature and the reaction was stopped by the addition of AEBSF ($0.2\ \text{mg ml}^{-1}$ final concentration). The cells were then centrifuged and supernatant removed. The pellets were then re-suspended in LDS loading buffer, boiled for 10 min, centrifuged for 10 min, and then separated on a NuPAGE Novex 4–12% Bis-Tris gel. The samples were then transferred to a PVDF membrane using an iBlot system (Invitrogen) and western blot analysis performed using a polyclonal anti-TbpA antibody (1:1,000) and a monoclonal anti-His HRP-conjugated antibody (Sigma) (1:5,000). Here, each membrane was blocked with 2% BSA in $1\times$ PBST for 15 min and then probed with either the anti-TbpA or anti-His HRP conjugated antibody for 30 min. The anti-TbpA membrane was then washed $2\times$ with 2% BSA in $1\times$ PBST and then probed with anti-mouse HRP-conjugated secondary antibodies for an additional 30 min. Both membranes were then washed $2\times$ with $1\times$ PBST, $2\times$ with $1\times$ PBS, and then imaged using the colorimetric substrate 3,3'-diaminobenzidine (Sigma). The results from the mutants were compared to wild-type TbpA to determine which constructs were being presented on the surface of the cells.

Transferrin competition assays. To determine if the affinity of hTF to either TbpA or TbpB is affected by the conformation or coordination state of the N lobe, we performed an ELISA competition assay using apo-hTF, holo-hTF, hTF–Fe_N (iron bound in N lobe only) and hTF–Fe_C (iron bound in C lobe only), which were expressed in BHK cells and purified as described²³. Here, His–TbpA (20 ng) or His–TbpB (20 ng) was incubated for 15 min in a final volume of $100\ \mu\text{l}$ in a 96-well Ni-NTA HiSorb plate (Qiagen) in $1\times$ PBS (0.05% cymal-6 was added to all buffers for His–TbpA). Wells were washed $2\times$ with PBST, blocked with 2% BSA in PBST for 15 min, and then incubated with each of the transferrin constructs (apo, holo, Fe_N, Fe_C) (100 ng each) for 15 min. Wells were washed again $2\times$ with PBST and then probed for 20 min with HRP-conjugated hTF (1:1,000) (Jackson ImmunoResearch) in $100\ \mu\text{l}$ final volume. Wells were washed again $2\times$ with PBST and then $2\times$ with PBS and imaged using $100\ \mu\text{l}$ 3,3',5,5'-tetramethylbenzidine substrate (Sigma) for ~ 5 min and terminated using Stop solution (Sigma). Data were collected and analysed as described above.

Small-angle X-ray scattering analysis. The TbpB–hTF complex was dialysed overnight at 4°C into TBS, pH 7.4 (25 mM Tris, 137 mM NaCl, 3 mM KCl) and

then filtered using a $0.2\ \mu\text{m}$ spin filter. Data were collected at concentrations of 1, 2.5 and $5\ \text{mg ml}^{-1}$ at Stanford Synchrotron Radiation Lightsource beamline BL4-2. Data reduction and analysis were performed using the beamline software SASAtool. The program AutoGNOM of the ATSAS suite⁴¹ was used to generate $P(r)$ curves and to determine maximum dimension (D_{max}) and radius of gyration (R_g) from the scattering intensity curve ($I(q)$ versus q) in an automatic, unbiased manner, and rounds of manual fitting in GNOM⁴² were used to verify these values. *Ab initio* molecular envelopes were computed by the programs GASBOR²⁴. Ten iterations of GASBOR were averaged using DAMAVER⁴³. Docking of the TbpB and diferric hTF crystal structures into the molecular envelope was performed manually, guided by both previous docking studies²⁰ and mutagenesis results. Figures were made with PyMOL and annotated and finalized with Adobe Illustrator.

Modelling the TbpA–TbpB–hTF triple complex. The *in silico* TbpA–TbpB–hTF triple complex was assembled based on our crystal structures (TbpA–(apo)hTF, diferric hTF, TbpB) and SAXS analysis (TbpB–(holo)hTF) reported here. The crystal structure (TbpA–hTF) was aligned with our TbpB–hTF model using the C1 subdomain of hTF as a reference, yielding a triple complex containing a 1:1:1 ratio of TbpA, TbpB and hTF. Figures were made with PyMOL and/or Maya (Autodesk) and annotated and finalized with Adobe Illustrator.

Electron microscopic analysis. The triple complex (TbpA–TbpB–(holo)hTF) was prepared from separately purified components by first forming a complex between TbpB and (holo)hTF, which was purified by size-exclusion chromatography in $1\times$ PBS. Cymal-6 was added to a final concentration of 0.05% and purified TbpA (in $1\times$ PBS, 0.05% Cymal 6) was added to the mixture using an excess of the TbpB–(holo)hTF complex. The triple complex (which retains a 1:1:1 stoichiometry) was isolated by sized exclusion chromatography in buffer A ($1\times$ PBS in 0.05% Cymal 6) and used immediately for EM experiments. The complex was diluted with buffer A to an optimal concentration for EM (determined empirically to be $\sim 1\ \mu\text{g ml}^{-1}$). Drops ($4\ \mu\text{l}$ each) were applied to carbon-coated, glow-discharged EM grids (EMS). After 1 min, the grid was blotted, washed twice with buffer A, once with distilled water, and then stained with 2% uranyl acetate. Grids were observed with a CM120 La-B6 electron microscope (FEI), operating at 120 kV. Micrographs were recorded on SO163 film (Kodak) at a nominal magnification of 45,000, and digitized on a Nikon Coolscan 9000 at a rate corresponding to $1.55\ \text{\AA}$ per pixel. The large majority of particles distributed evenly on the grid and were essentially uniform in size (~ 90 – $110\ \text{\AA}$ in diameter), indicative of a homogeneous population.

The particles were variable in substructure, suggesting that the molecules deposit on the grid in a variety of orientations. Accordingly, a data set of 4,240 particles was subjected to a 'reference free' classification, using SPIDER⁴⁴, EMAN⁴⁵ and Bsoft⁴⁶. Images were picked using a 256×256 pixel box, and binned four times (to $6.2\ \text{\AA}$ per pixel) to increase the signal-to-noise ratio and the speed of calculation. Initial reference-free classification and averaging were performed using EMAN; further classification was done in SPIDER, using principal component analysis (PCA) with three cycles of iteration. We chose to obtain 56 final class averages, based on a cluster distribution obtained from PCA.

The coordinates of the modelled triple complex were converted to a density map and low-pass filtered to $15\ \text{\AA}$. The sampling rate of the density map was set to be same as the EM images and two-dimensional projections were calculated at angular increments of 30° (Supplementary Fig. 14b). Comparisons and matching between the EM class averages and the model re-projections were done in terms of cross-correlation coefficients. A few ambiguous cases were reassigned by visual assessment. Figures were made with Chimera and annotated and finalized with Adobe Illustrator.

Molecular dynamics simulations. For simulations of TbpA bound to apo-hTF, a membrane-water system containing the protein complex was first built using VMD⁴⁷. The complex was placed in a DMPE bilayer as used previously⁴⁸, with the barrel of TbpA aligned with the membrane's hydrophobic core, and then fully solvated. Disulphide bonds for three pairs in TbpA and 19 pairs in hTF were added based on S–S proximity. Ca²⁺ and Cl[−] ions were added to a concentration of 100 mM, resulting in an initial size of 264,000 atoms. The system was equilibrated in stages for 13.5 ns, including 10 ns of fully unrestrained dynamics. The simulations were run using NAMD 2.7³³ in the NPT ensemble at a temperature of 310 K and a pressure of 1 atm; after the first 3.5 ns of equilibration, the area of the membrane was fixed. Other simulation parameters were set identically to those used previously⁴⁸. For steered molecular dynamics (SMD) simulations, the C α atom of the TbpA N-terminal plug domain residue was pulled in the $-z$ direction, away from the membrane, at a constant velocity of $5\ \text{\AA ns}^{-1}$ (refs 33, 49). To counterbalance the pulling forces, six residues at the extracellular periphery of the barrel domain were restrained in the z direction. An adaptive procedure was used to limit the maximum required system size during SMD simulations⁵⁰. When the extension of the unfolded region of the plug domain brought it near to the periodic boundary, the simulation was stopped, the unfolded region of the plug

domain distant from the barrel and membrane was cleaved, and the simulation restarted after a short equilibration of the water with the new N-terminal residue being pulled. With this procedure, used three times, approximately 150 Å of pulling was accomplished while keeping the system sizes below 300,000 atoms.

Sequence analysis and alignments. Sequence analysis and alignments were performed and analysed with the programs STRAP⁵¹ and JalView⁵². Figures were annotated and finalized with Adobe Illustrator.

34. Otwinowski, Z. & Minor, W. Processing of X-ray data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
36. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
38. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
39. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
40. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Petoukhov, M. V. & Svergun, D. I. Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Opin. Struct. Biol.* **17**, 562–571 (2007).
42. Svergun, O. & Genkina, O. A. The dependence of the dynamics of the extinction of a temporary connection on the recognizability of a reinforcing stimulus [in Russian]. *Zh. Vyssh. Nerv. Deiat. Im. I P Pavlova* **41**, 700–707 (1991).
43. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860–864 (2003).
44. Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3**, 1941–1974 (2008).
45. Ludtke, S. J. 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods Mol. Biol.* **673**, 157–173 (2010).
46. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
47. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
48. Gumbart, J., Wiener, M. C. & Tajkhorshid, E. Coupling of calcium and substrate binding through loop alignment in the outer-membrane transporter BtuB. *J. Mol. Biol.* **393**, 1129–1142 (2009).
49. Sotomayor, M. & Schulten, K. Single-molecule experiments *in vitro* and *in silico*. *Science* **316**, 1144–1148 (2007).
50. Gumbart, J., Wiener, M. C. & Tajkhorshid, E. Mechanics of force propagation in TonB-dependent outer membrane transport. *Biophys. J.* **93**, 496–504 (2007).
51. Gille, C. & Frommel, C. STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* **17**, 377–378 (2001).
52. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

Gain control by layer six in cortical circuits of vision

Shawn R. Olsen^{1*}, Dante S. Bortone^{1*}, Hillel Adesnik¹ & Massimo Scanziani¹

After entering the cerebral cortex, sensory information spreads through six different horizontal neuronal layers that are interconnected by vertical axonal projections. It is believed that through these projections layers can influence each other's response to sensory stimuli, but the specific role that each layer has in cortical processing is still poorly understood. Here we show that layer six in the primary visual cortex of the mouse has a crucial role in controlling the gain of visually evoked activity in neurons of the upper layers without changing their tuning to orientation. This gain modulation results from the coordinated action of layer six intracortical projections to superficial layers and deep projections to the thalamus, with a substantial role of the intracortical circuit. This study establishes layer six as a major mediator of cortical gain modulation and suggests that it could be a node through which convergent inputs from several brain areas can regulate the earliest steps of cortical visual processing.

Primary sensory areas in the cerebral cortex are composed of a stack of six neuronal layers¹. Anatomical and physiological data indicate that these layers are interconnected through vertical excitatory axons^{2–6}, suggesting that sensory processing in any given layer may be modulated by activity in several other layers. However, so far the exact contribution of each layer to cortical processing is unclear.

Here we address the role of layer six (L6) in mouse visual cortex, whose excitatory neurons not only project to more superficial layers but also to the primary sensory thalamic nuclei^{3,7–11}, the main source of sensory input to the cortex (Fig. 1a). L6 may thus influence cortical sensory responses directly through intracortical projections and indirectly through corticothalamic projections. Corticothalamic projections were reported to be both suppressive and facilitatory on thalamic activity, depending on the precise alignment between L6 and thalamic neurons (for reviews see refs 12–16). By contrast, how sensory responses in cortex are affected by L6 activity has remained largely unexplored^{17,18}. Furthermore, the relative contribution of intracortical versus corticothalamic projections in modulating cortical responses is currently unknown. The paucity of information is due to the lack of experimental tools for selectively manipulating activity in L6 without directly perturbing other cortical layers.

L6 neurons of the *Ntsr1*-Cre GN220 line

To control the activity of L6 we took advantage of a Cre-recombinase Bac transgenic mouse line that is reported to selectively label L6 neurons (NTSR1-Cre GN220)¹⁹. In the forebrain of these mice Cre expression was restricted to excitatory L6 neurons of the cerebral cortex (Fig. 1b and Supplementary Fig. 1). In primary visual cortex (V1) these neurons represented ~65% of the L6 excitatory neuronal population and, consistent with classification of L6 neurons in this region¹, could be subdivided into two morphologically distinct categories: those whose apical dendrites ended in L4 and those that extended to L1 (Fig. 1b and Supplementary Fig. 1g, h). Furthermore, consistent with the corticothalamic projections originating from L6 in V1 (ref. 8), Cre-expressing neurons projected to the dorsolateral geniculate nucleus (dLGN; the primary thalamic visual nucleus) and the nucleus reticularis thalami (NRT; the main thalamic inhibitory nucleus) (Fig. 1b and

Supplementary Fig. 1d, e). Thirty-five percent of L6 excitatory neurons in V1 did not express Cre and these were morphologically distinct from the Cre-expressing population (Supplementary Fig. 1g).

To manipulate the activity of L6 neurons we conditionally expressed the light-sensitive cation channel channelrhodopsin 2 (ChR2)^{20,21} in V1 using viral injection into NTSR1-Cre mice (Supplementary Fig. 2a). A linear multichannel probe recorded the spiking activity of neurons located across the vertical depth of cortex. Light-emitting diode (LED) illumination of the cortical surface for 500 ms with blue light (470 nm) increased the activity of L6 neurons in V1 of anesthetized animals (Fig. 1c–e and Supplementary Fig. 2b). This increase was not due to direct stimulation of the retina by the LED as it was absent in uninjected animals (Supplementary Fig. 2g).

L6 activity suppresses other layers

To determine how L6 activation affects visual responses in other layers, we presented drifting gratings, and alternated control trials (visual stimulus only) with trials in which L6 was photostimulated (Fig. 1c). Notably, photostimulation of L6 rapidly and reversibly suppressed visually evoked multi-unit activity throughout the depth of the cortex (Fig. 1d). L6 photostimulation also reduced spontaneous activity (Supplementary Fig. 3d, e). This effect was absent in uninjected animals (Supplementary Fig. 2g). The suppressive action of L6 was similar across L2/3, L4 and L5 and was monotonic (Fig. 1e,f): that is, increasing L6 activity by increasing illumination intensity progressively suppressed visual responses, eventually abolishing nearly all evoked activity (strongest illumination reduced activity by $81 \pm 5\%$, $84 \pm 3\%$, and $84 \pm 3\%$ for L2/3, L4 and L5, respectively; $P < 10^{-5}$). Because multi-unit activity is dominated by neurons with high firing frequencies, we determined the effect of L6 photostimulation on isolated single units whose average visually evoked firing rate varied over a 20-fold range. Isolated units were suppressed by L6 photostimulation (Fig. 1g), irrespective of their firing rates (Fig. 1h; 91.1% of units were suppressed and 7.8% were facilitated, and all facilitated units were fast-spiking, putative inhibitory cells (Supplementary Fig. 4a–d). Furthermore, in the same way as for multi-unit activity, L6 photostimulation monotonically suppressed single units (Fig. 1i, j; strongest

¹Howard Hughes Medical Institute, Center for Neural Circuits and Behavior, Neurobiology Section and Department of Neuroscience, University of California San Diego, La Jolla, California 92093-0634, USA.

*These authors contributed equally to this work.

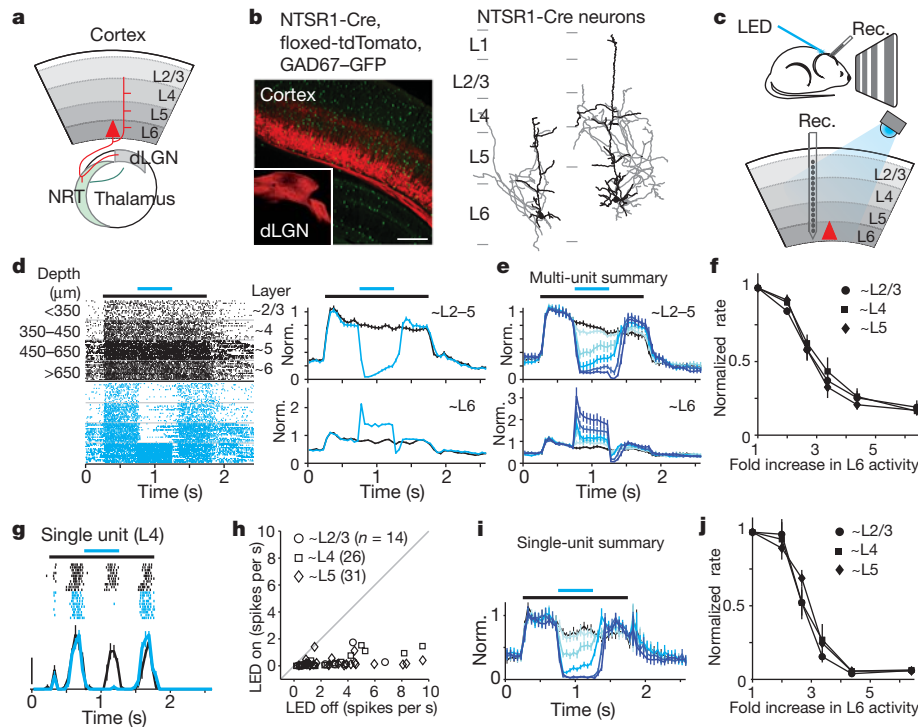


Figure 1 | Photostimulation of L6 suppresses visual responses in the other layers. **a**, Schematic of L6 projections. Red triangle represents an L6 pyramidal neuron. **b**, Left, coronal section of V1 from an NTSR1-Cre, floxed-tdTomato, GAD67-GFP mouse. Inset, L6 projection to dLGN (V1 of NTSR1-Cre mouse was injected with floxed-tdTomato virus). Scale bar, 250 μ m (125 μ m for inset). Right, the two types of L6 neurons that are labelled by the NTSR1-Cre line. Black, dendrites; grey, axons. **c**, Schematic of experimental setup. Rec., recording probe. **d**, Cortical visual responses with (blue) and without (black) L6 photostimulation. Left, raster plot of multi-unit activity grouped by depth. Control and photostimulation trials were interleaved but are separated here for

clarity. Black bar, visual stimulus (1.5 s); blue bar, LED illumination (0.5 s). Right, normalized (Norm.) peristimulus time histogram (PSTH); top, upper layers; bottom, L6. **e**, Summary ($n = 6$ experiments). The control is shown in black and increasing LED intensities in darker blues. **f**, Suppression of multi-unit activity with increasing L6 activity. **g**, Visual response of a single L4 unit with (blue) and without (black) L6 photostimulation. Scale bar, 20 spikes per s. **h**, Response of each regular spiking unit with and without strong photostimulation of L6. **i**, Average normalized PSTH ($n = 47$ units tested with 5 LED intensities). Colours are the same as in **e**. **j**, Suppression of single-unit activity. Error bars, mean \pm s.e.m.

illumination reduced activity by $91 \pm 4\%$, $93 \pm 2\%$, and $92 \pm 2\%$ for L2/3, L4 and L5, respectively; $P < 10^{-9}$). Thus, these data show that stimulation of L6 excitatory neurons suppresses visually evoked responses in L2/3, L4 and L5 of V1.

L6 activity does not affect tuning

Like in other mammals, neurons in mouse V1 differentially respond to gratings of different orientations^{22,23}. We determined whether L6 stimulation affects the orientation tuning of V1 neurons. We generated tuning curves by presenting gratings drifting in 8–12 different directions and alternated control trials with trials in which L6 was photostimulated (Fig. 2a, b). We used a low LED intensity to suppress cortical visual responses partially, and considered units that were suppressed by between 10% and 75% (average suppression $42 \pm 3\%$, $n = 55$). Tuning curves of individual, isolated units were averaged into a population tuning curve (Fig. 2b, d; see methods). Remarkably, photostimulation of L6 resulted in the precise scaling of the tuning curve; that is, it reduced visually evoked responses by a similar fraction irrespective of presented orientation. This is clearly illustrated by plotting the normalized firing rates of the population tuning curve under control versus L6 photostimulation conditions (Fig. 2e). The data points fit well with a line whose slope is 0.56 and intercepts the y axis close to the origin. Thus, photostimulation of L6 did not affect preferred orientation, tuning width or the orientation selectivity index (OSI) of cortical neurons throughout L2/3, L4 and L5 (Fig. 2c; for L2/3, L4 and L5, respectively, the mean change in preferred orientation was $3 \pm 3^\circ$ ($P = 0.22$), $0 \pm 5^\circ$ ($P = 0.9$) and $-4 \pm 5^\circ$ ($P = 0.48$), mean change in tuning width was $-1 \pm 4^\circ$ ($P = 0.8$), $6 \pm 4^\circ$ ($P = 0.15$) and $-6 \pm 6^\circ$ ($P = 0.3$), and mean change in OSI was -0.09 ± 0.07 ($P = 0.23$),

0.7 ± 0.04 ($P = 0.14$), -0.06 ± 0.05 ($P = 0.22$)). L6 photostimulation also resulted in a scaling of V1 responses to stimuli of increasing contrast (the contrast response function; Supplementary Fig. 5b). These data demonstrate that in primary visual cortex L6 selectively controls the gain of cortical responses to visual stimuli.

A potential concern in stimulating L6 with ChR2 is that the spatially uniform activation and the temporal pattern generated in L6 neurons may differ from visually evoked activity patterns, and thus the physiological activity of L6 neurons and L6 photostimulation may affect cortical activity in different ways. Furthermore, anaesthesia may change the impact of L6 on cortical responses to sensory stimuli. To address these issues, we optogenetically suppressed visually evoked activity in L6 in awake animals and determined the resulting effect on more superficial layers (Supplementary Fig. 6). Animals were head fixed but otherwise kept unrestrained on a passive circular treadmill (see Methods). L6 activity was suppressed using conditionally expressed light-sensitive hyperpolarizing opsins archaeorhodopsin²⁴ and halorhodopsin 3.0 (NpHR3.0) (ref. 25). LED illumination with amber light (590 nm), although reducing visually evoked L6 activity by $\sim 30\%$ (Supplementary Fig. 6e), significantly facilitated visual responses of isolated units throughout the other layers (Fig. 2f, g and Supplementary Fig. 6). The facilitation was not due to direct LED illumination of the retina, as it was absent in uninjected animals (Supplementary Fig. 6f). Thus, suppression of L6 facilitates visually evoked activity in L2/3, L4 and L5, indicating that even physiologically generated L6 activity exerts a suppressive action onto these layers. Furthermore, suppression of L6 resulted in the precise scaling of the tuning curve (for the tuning curve analysis we considered units that were facilitated by at least 10% (average facilitation $41 \pm 7\%$, $n = 52$)).

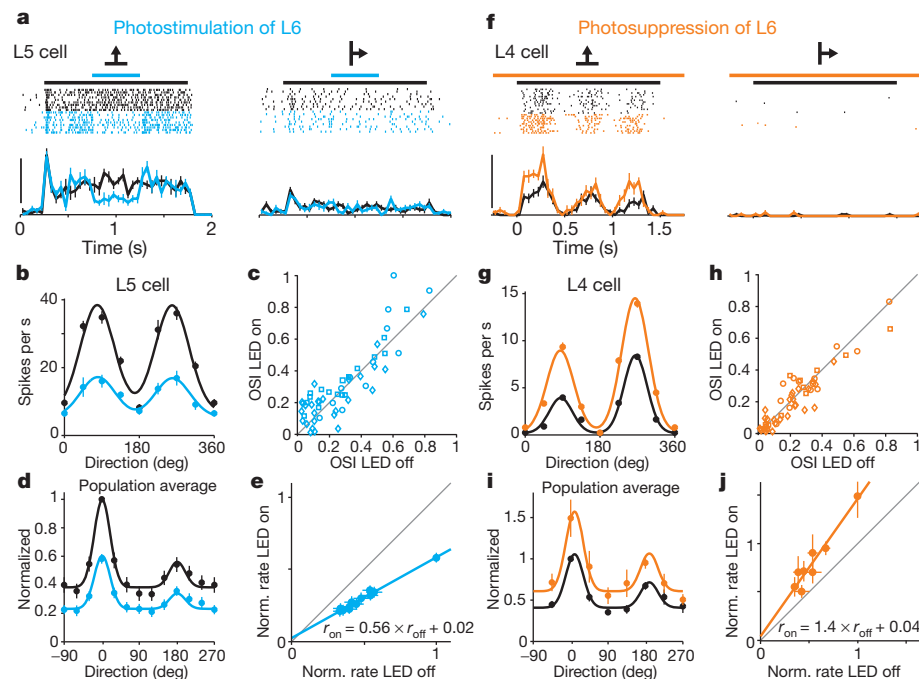


Figure 2 | L6 bidirectionally modulates the gain of visual responses without altering tuning. **a**, Visual responses of an L5 neuron with (blue) and without (black) L6 photostimulation. Raster plots and peristimulus time histograms for two out of eight tested visual stimulus directions. Scale bar, 40 spikes per s. **b**, Tuning curves for the neuron in **(a)**. **c**, The OSI for each neuron with and without photostimulation of L6. **d**, Population tuning curve with (blue) and without (black) L6 photostimulation ($n = 55$). Black curve, fit using the sum of two Gaussians; blue curve, black curve scaled by the slope of linear fit in **e**. **e**, Control response plotted against response with L6 photostimulation (data from **c**). Blue, linear fit ($r^2 = 0.98$). Black bar, visual stimulus (1.5 s); blue bar,

LED illumination (0.5 s). **f**, Visual response of an L4 neuron with (orange) and without (black) L6 photosuppression. Scale bar, 50 spikes per s. Orange bars, illumination with an amber-coloured LED (1.95 s); black bar, visual stimulation (1.5 s). **g**, Tuning curves for neuron in **(f)**. **h**, OSI for each isolated unit with and without photosuppression of L6. **i**, Population tuning curves with and without L6 photosuppression ($n = 52$). Black curve, fit using sum of two Gaussians; orange curve, black curve scaled by slope of linear fit in **j**. **j**, Control response plotted against response with L6 photostimulation (data from **i**). Orange, linear fit ($r^2 = 0.92$). Error bars, mean \pm s.e.m.

The plot of normalized firing rates under control versus L6 photosuppression conditions was well fit by a line whose slope is 1.4 and intercepts the y axis very close to the origin (Fig. 2j). Consistent with this, suppressing L6 did not affect preferred orientation, tuning width or orientation selectivity (Fig. 2h; for L2/3, L4 and L5, respectively, the mean change in preferred orientation was $2 \pm 3^\circ$ ($P = 0.41$), $0 \pm 2^\circ$ ($P = 0.95$) and $-4 \pm 4^\circ$ ($P = 0.35$) degrees, mean change in tuning width was $-2 \pm 4^\circ$ ($P = 0.68$), $0 \pm 3^\circ$ ($P = 0.94$) and $-1 \pm 4^\circ$ ($P = 0.77$) degrees, and mean change in OSI was -0.01 ± 0.03 ($P = 0.22$), 0.02 ± 0.02 ($P = 0.50$) and -0.03 ± 0.03 ($P = 0.22$)). Taken together, these results demonstrate that visually driven L6 activity in awake animals controls the gain of cortical responses to visual stimuli.

L6 intracortical and subcortical pathways

Two pathways could potentially mediate the suppression exerted by L6 on cortical activity. On one hand, L6 neurons project to the thalamus, where they can influence visually generated activity before it even reaches the cortex. On the other hand, L6 neurons also project to more superficial layers where they could directly modulate cortical activity. We addressed the impact of both projections. We performed extracellular recordings from the dLGN while photostimulating L6 in V1 (Fig. 3a). dLGN relay neurons were identified based on their visual response properties and characteristic firing pattern (Supplementary Fig. 7d). Photostimulation of L6 led to a rapid, reversible and monotonic reduction of visually evoked activity in dLGN relay neurons (Fig. 3b, c; strongest illumination: $76 \pm 4\%$ reduction; $P < 10^{-10}$, $n = 32$), without, however, markedly modifying their firing mode (burst prevalence: $12 \pm 6\%$ in control; $6 \pm 3\%$ after reducing dLGN activity by 30% with L6 photostimulation, $P = 0.08$; Supplementary Fig. 7e, f). This indicates that L6 stimulation suppresses

dLGN activity. To test whether visually evoked activity in L6 also suppresses dLGN activity we silenced the cortex optogenetically (by photostimulating parvalbumin-expressing inhibitory neurons in V1 with Chr2; see Methods and Supplementary Fig. 8). Consistent with the suppressive action of L6 stimulation on dLGN, silencing the cortex strongly facilitated dLGN activity (Fig. 3d–f; average facilitation $87 \pm 25\%$ ($P = 0.002$, $n = 18$)). *In vitro* recordings demonstrated that the suppressive action of L6 was due to the generation of disynaptic inhibition onto dLGN relay neurons, at least in part through the recruitment of NRT inhibitory neurons (and possibly through the recruitment of local inhibitory neurons in dLGN²⁶) (Supplementary Fig. 9). Thus, these results reveal that L6 can effectively suppress visual responses in the dLGN.

If L6 suppresses cortical visual responses indirectly, by suppressing the dLGN, this suppression should precede V1 suppression by a few milliseconds. We tested this prediction by performing simultaneous recordings from both dLGN and V1 and compared the onset of suppression in these two structures upon L6 photostimulation (Fig. 3g). Surprisingly, cortical suppression preceded dLGN suppression by a few milliseconds (Fig. 3h). This result suggests that L6 activity may suppress cortical visual responses through an alternative circuit. Because L6 neurons send axons to the upper layers of cortex we tested whether these projections can suppress cortical activity independently of the corticothalamic projections. For this, we performed *in vitro* whole-cell recordings from neurons in L2/3, L4, L5 and L6 in coronal slices of V1 (Fig. 4a); this slicing plane disconnects V1 from dLGN.

Photostimulation of L6 *in vitro* generated both excitatory and inhibitory postsynaptic currents (EPSCs and IPSCs, respectively) onto L2/3, L4, L5 and L6 pyramidal cells (L6 recordings included only those pyramidal cells not expressing Chr2) (Fig. 4b). IPSCs were of disynaptic (or polysynaptic) origin as they were entirely blocked by

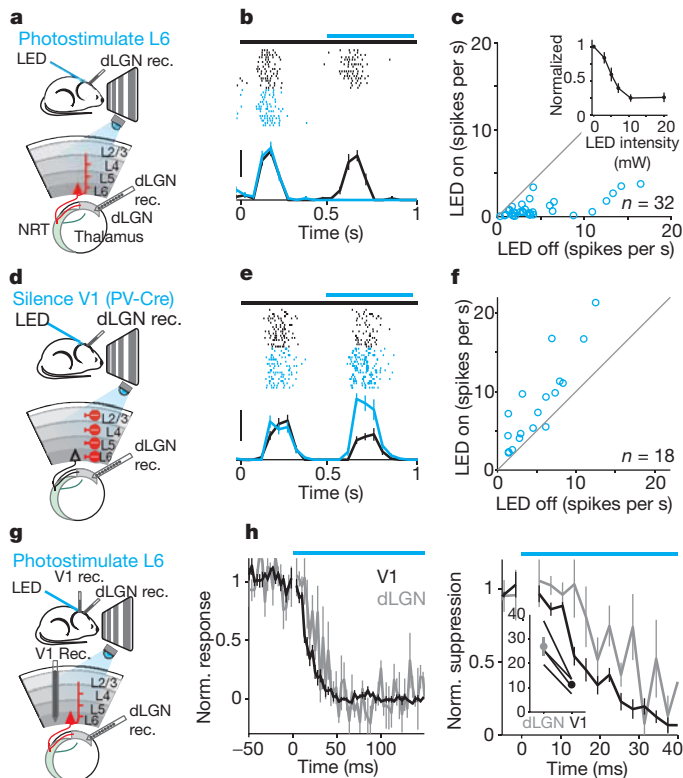


Figure 3 | Photostimulation of L6 suppresses cortex faster than it suppresses dLGN. **a**, Schematic of the experimental setup. **b**, Visual response of dLGN unit with (blue) and without (black) L6 photostimulation. Scale bar, 20 spikes per s. Black bar, visual stimulus (1 s); blue bar, LED illumination (0.5 s). **c**, Average response of each dLGN unit with and without L6 photostimulation. Inset, monotonic suppression of dLGN. **d**, Schematic of setup for silencing V1 by photostimulation of parvalbumin inhibitory neurons. **e**, Visual response of dLGN unit with and without photo-silencing of V1. Scale bar, 30 spikes per s. Black bar, visual stimulus (1 s); blue bar, LED illumination (0.5 s). **f**, Average response of each dLGN unit with and without cortical silencing. **g**, Schematic of experimental setup. **h**, Left, time-course of L6-mediated suppression of dLGN (grey) and V1 (black) ($n = 4$). Residual response during maximal suppression set to zero (see Methods). Bin size, 3 ms. Right, the same data on an expanded timescale. The first bin at LED onset was blanked to remove LED-induced artefact. Inset, time to suppression exceeding two standard deviations from baseline activity in dLGN and V1 for four experiments ($P = 0.012$). Error bars, mean \pm s.e.m. Inset, y-axis units are ms.

glutamatergic antagonists (Supplementary Fig. 10b). Furthermore, the activity pattern generated by L6 photostimulation was similar to the activity pattern generated *in vivo* (Supplementary Fig. 2b, h). IPSCs were larger than EPSCs, despite the fact that both currents were recorded with a similar driving force (IPSCs were recorded near the reversal potential for EPSCs and vice versa). Indeed, excitatory charge accounted for only 10% or less of the total charge, depending on the layer (Fig. 4c) or sublayer (Supplementary Fig. 10c, d). These results show that V1 contains the necessary circuitry for L6 to generate disynaptic inhibition onto L2/3, L4, L5 and onto itself.

To determine whether L6 can suppress neuronal spiking across L2/3, L4, L5 and L6 through these disynaptic IPSCs, we performed current-clamp recordings in the perforated patch configuration (to preserve the intracellular ionic composition) and triggered spiking by injecting depolarizing current pulses. Photostimulation of L6 significantly suppressed firing of pyramidal cells across all layers (Fig. 4d; firing rate was reduced by $48 \pm 10\%$, $84 \pm 7\%$, $55 \pm 19\%$ and $75 \pm 11\%$ for L2/3, L4, L5 and L6, respectively; $P \leq 0.01$). To rule out the possibility that this suppression was a result of uniformly activating large portions of L6 we restricted the area of activation to a small spot of approximately $100 \mu\text{m}$ in diameter while recording

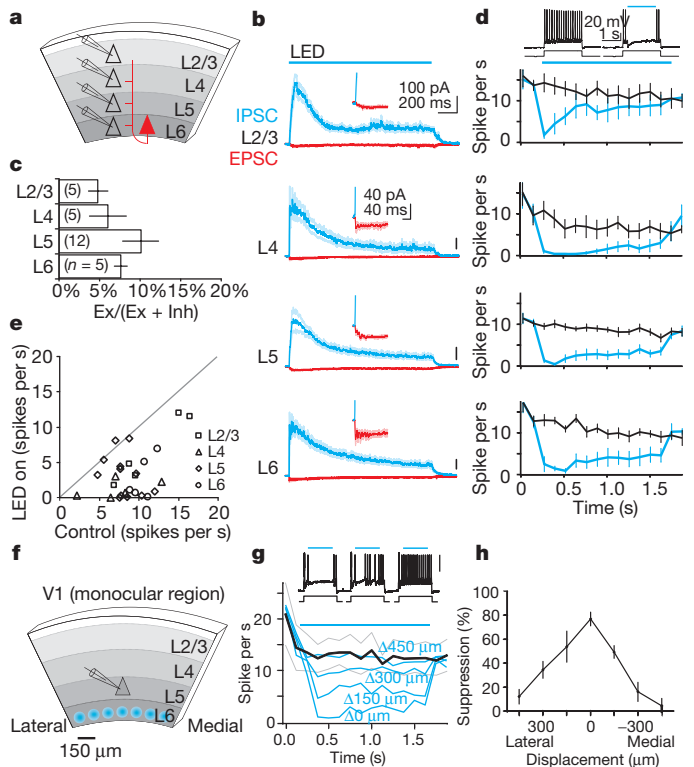


Figure 4 | Photostimulation of L6 recruits intracortical synaptic inhibition. **a**, Schematic of *in vitro* experimental setup. **b**, Average IPSCs (blue) and EPSCs (red) recorded in pyramidal cells during photostimulation of L6. Synaptic currents are averages of $n = 5-12$ cells. Inset, onset of EPSC. **c**, Histogram of excitatory charge as a percentage of total charge. Ex, excitation; Inh, inhibition. **d**, Traces show perforated patch recording from L5 pyramidal cell in response to depolarizing current injection with (right) and without (left) L6 photostimulation. Graphs, spike rate with and without L6 photostimulation. **e**, Average spike rate in control versus spike rate with L6 photostimulation for each cell. **f**, Schematic of experimental setup for focal photostimulation. **g**, Traces, spiking of L5 pyramidal cell to depolarizing current injection with focal photostimulation of L6 at three progressively more distant positions (left to right). Graph shows spike rate in control (black) and with focal photostimulation of L6 (blue) ($n = 4$). Delta indicates the medial or lateral distance from the radial axis through the recording site. **h**, Percentage of spike suppression plotted against horizontal displacement. Error bars, mean \pm s.e.m.

from a L5 neuron (Fig. 4f). Even when activating a restricted area of L6, the firing of L5 neurons was robustly suppressed (Fig. 4g). The suppression was maximal when L6 photostimulation was aligned with the recorded L5 neuron along the cortical radial axis, and decreased progressively as the photostimulation spot was translated tangentially (Fig. 4g, h). These results demonstrate that V1 can efficiently suppress activity in L2/3, L4, L5 and L6 in the absence of thalamus.

Major role of L6 intracortical circuits

Taken together, these results indicate that L6 can modulate cortical responses to visual stimuli through two independent circuits: indirectly, through the corticothalamic circuit and directly, through the intracortical circuit. To test whether one of these two circuits has a dominant role, we examined how much of the V1 suppression is predicted by dLGN suppression. We first established the transfer function between dLGN and V1. For this we performed simultaneous *in vivo* recordings from these two structures while presenting full-field drifting gratings of varying contrasts to obtain contrast response functions for the dLGN and V1 (Fig. 5a, b). By plotting dLGN versus V1 activity at each contrast we obtained the transfer function from dLGN to V1, which provides the response of V1 to various levels of dLGN activity (Fig. 5c). Finally, we presented the strongest contrast and photostimulated L6 to reduce dLGN activity while simultaneously monitoring V1 activity.

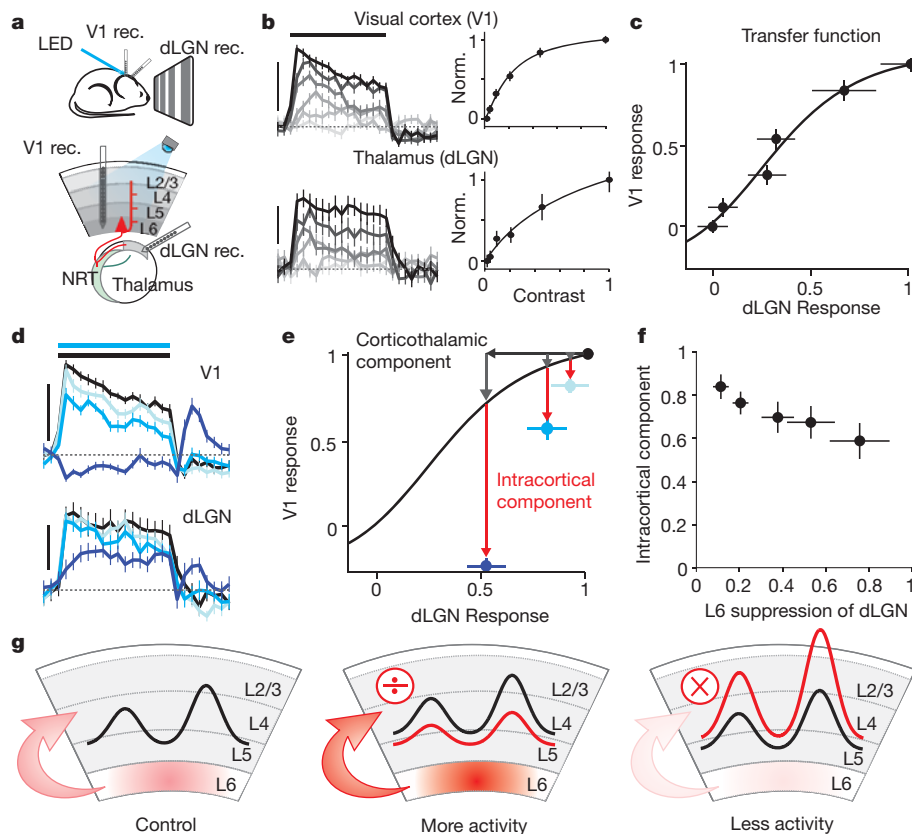


Figure 5 | L6 suppresses upper layers largely through intracortical circuits. **a**, Schematic of experimental setup. **b**, Simultaneously recorded multi-unit responses to increasing contrasts (light to dark) in V1 (top) and dLGN (bottom). All spikes that were recorded above L6 ($\leq 650 \mu\text{m}$) were included in V1 multi-unit activity. Scale bar, 200 spikes per s for V1; 100 spikes per s for dLGN. Black bar, visual stimulus (1.5 s). Dotted line, baseline activity. Right, contrast-response functions. **c**, dLGN–V1 transfer function derived by plotting normalized response in V1 versus dLGN (from **b**). Fit, hyperbolic ratio function. **d**, Simultaneously recorded multi-unit responses to maximal contrasts in V1 and dLGN without photostimulation (black) or while photostimulating L6 with increasing LED intensities (progressively darker blue). Same experiment as in **b** and **c**. Black bar, visual stimulus (1.5 s); blue bar, LED illumination (1.5 s). Scale bars are the same as in **b**. **e**, V1 versus dLGN response to maximal contrast under control condition (black data point) or during three progressively stronger photostimulations of L6 (light, medium and dark blue, data from **d**). V1 responses are suppressed more than predicted by transfer function (red arrows) even for photostimulations that reduce dLGN activity only $\sim 10\%$ (light blue). **f**, Average intracortical component of suppression as a function dLGN suppression ($n = 5$ experiments). Intracortical component (red arrow in **e**) is quantified as a fraction of total V1 suppression (grey arrows plus red arrows in **e**). **g**, Schematic of the main finding. Error bars, mean \pm s.e.m.

We reasoned that if the ensuing reduction of V1 activity matches the reduction predicted by the transfer function, the modulation of cortical responses by L6 is mainly due to dLGN suppression by the corticothalamic circuit. However, if the reduction in V1 activity exceeds that predicted by the transfer function, the additional reduction can be attributed to the intracortical circuit. We reduced dLGN activity by $\sim 10\%$, 20% and 50% through activation of L6 with three progressively stronger illuminations (Fig. 5d). Notably, even the smallest reduction in dLGN activity (10%) was accompanied by a reduction in V1 activity that largely exceeded that predicted by the transfer function (Fig. 5e). Furthermore, a 50% suppression of dLGN activity was accompanied by a complete suppression of visually evoked activity in V1. In this experiment a large fraction of V1 suppression (73% averaged over five LED levels) exceeded the transfer function prediction and must therefore be attributed to the intracortical circuit (average intracortical component over all experiments $73 \pm 5\%$, $n = 5$; Fig. 5f). Furthermore, given the relatively minor effects on the prevalence of burst firing in dLGN neurons (Supplementary Fig. 7f), this effect cannot be attributed to a change in the firing pattern of dLGN neurons. These results indicate that L6 suppresses cortical responses to visual stimuli mainly through intracortical circuits.

Discussion

Taken together, this study shows that L6 modulates visually evoked activity across L2/3, L4 and L5. This modulation occurs continuously through visually driven L6 activity, as shown in awake animals, and does not affect orientation tuning indicating that L6 selectively controls the gain of cortical visual responses. Finally, despite suppression of the dLGN, cortical gain control by L6 is executed largely by intracortical circuits.

Response gain modulation is a fundamental cortical operation²⁷ that is crucially involved in sensory representation and sensorimotor integration. For example, visual responses in parietal cortex are gain modulated by gaze direction²⁸. Furthermore, gain modulation may underlie

the effects of attention on cortical responses to visual stimuli^{29,30}. However, the neuronal circuits that implement this operation have remained largely unknown. Identifying L6 as a contributor to cortical response gain modulation is an important step in dissecting the specific functions of distinct circuits in cortical processing. The suppressive action of L6 that is described here markedly differs from the facilitatory impact of other layers on cortical activity^{3,31,32} (for example, L2/3 facilitates L5 (ref.³²)) and points towards a very distinct function of different layers in sensory processing. The cortical GABAergic interneuron subtype (or subtypes)^{33,34} that is recruited by L6 activity and mediates the reported suppressive effect remains to be identified, but may include fast spiking neurons (Supplementary Fig. 4). Although the exact synaptic mechanisms underlying gain control by L6 remains to be elucidated, either a proportional change in excitation and inhibition^{35,36} or the modulation of only one of the two opposing conductances³⁷ may underlie the operation. The columnar organization of L6 pyramidal cell projections to more superficial layers¹⁰ ensures that L6-mediated suppression is restricted to the cortical domains that are directly above the activated L6 region (Fig. 4g, h). This topographic organization could allow the cortex to differentially modulate the gain of V1 responses to stimuli located in distinct regions of visual space.

L6 has been suggested to contribute to ‘end inhibition’, the suppression of cortical responses by bars above a given size¹⁷. The powerful inhibitory currents generated by L6 onto more superficial pyramidal cells may represent the underlying cellular mechanism.

Previous studies addressing the role of corticothalamic feedback projections through focal pharmacological perturbation of L6 neurons have typically reported a facilitation of functionally or topographically aligned thalamic neurons overlaid by broader surround suppression¹⁶, resulting in changes to both spatial and temporal tuning properties of these neurons^{15,38–41}. Our data obtained using full-field visual stimulation are consistent with this model, in which spatial summation of individual inhibitory surrounds will result in a net suppressive effect of the corticothalamic feedback projection. Future studies combining

optogenetic approaches with focal stimulation of visual space will reveal how fine-scale corticothalamic circuits^{39,42} interact with intra-cortical L6 circuits to influence visual processing in the cortex.

L6 in V1 receives convergent inputs from a variety of brain regions, including higher cortical areas⁴³ as well as thalamus¹¹. These various brain regions could thus influence, through L6, the gain of visual responses during the very initial steps of visual cortical processing.

METHODS SUMMARY

ChR2, archaerhodopsin and NpHR3.0 were conditionally expressed in mouse V1 via stereotactic injection of adenoassociated viruses into NTSR1-Cre mice¹⁹. For recordings under anaesthesia, mice were injected with 5 mg kg⁻¹ chlorprothixene and 1.2 g kg⁻¹ urethane. For awake experiments, a craniotomy was performed under isoflurane anaesthesia (1–1.5%), and then a previously implanted head-plate was used to fix the mouse on a circular treadmill and the anaesthetic was removed. *In vivo* extracellular recordings were made from V1 and dLGN using multichannel silicon probes. Visual stimuli were displayed on an LCD screen. Microbial opsins were photoactivated using a blue (470-nm) or amber (590-nm) LED placed above the thinned skull. *In vitro* whole-cell recordings were performed as previously described³².

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 August 2011; accepted 4 January 2012.

Published online 22 February 2012.

- Lorente de No, R. in *Physiology of the Nervous System* (ed. Fulton, J.F.) 274–301 (Oxford Univ. Press, 1943).
- Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004).
- Lefort, S., Tómm, C., Floyd Sarria, J. C. & Petersen, C. C. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* **61**, 301–316 (2009).
- Thomson, A. M. & Bannister, A. P. Interlaminar connections in the neocortex. *Cereb. Cortex* **13**, 5–14 (2003).
- Callaway, E. M. Local circuits in primary visual cortex of the macaque monkey. *Annu. Rev. Neurosci.* **21**, 47–74 (1998).
- Dantzker, J. L. & Callaway, E. M. Laminar sources of synaptic input to cortical inhibitory interneurons and pyramidal neurons. *Nature Neurosci.* **3**, 701–707 (2000).
- Thomson, A. M. Neocortical layer 6, a review. *Front. Neuroanat.* **4**, 13 (2010).
- Bourassa, J. & Deschenes, M. Corticothalamic projections from the primary visual cortex in rats: a single fiber study using biocytin as an anterograde tracer. *Neuroscience* **66**, 253–263 (1995).
- Binzegger, T., Douglas, R. J. & Martin, K. A. Stereotypical bouton clustering of individual neurons in cat primary visual cortex. *J. Neurosci.* **27**, 12242–12254 (2007).
- Zhang, Z. W. & Deschenes, M. Intracortical axonal projections of lamina VI cells of the primary somatosensory cortex in the rat: a single-cell labeling study. *J. Neurosci.* **17**, 6365–6379 (1997).
- Jones, E. G. *The Thalamus* (Cambridge Univ. Press, 2007).
- Guillery, R. W. & Sherman, S. M. Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron* **33**, 163–175 (2002).
- Sillito, A. M. & Jones, H. E. Corticothalamic interactions in the transfer of visual information. *Phil. Trans. R. Soc. Lond. B* **357**, 1739–1752 (2002).
- Briggs, F. & Usrey, W. M. Emerging views of corticothalamic function. *Curr. Opin. Neurobiol.* **18**, 403–407 (2008).
- Cudeiro, J. & Sillito, A. M. Looking back: corticothalamic feedback and early visual processing. *Trends Neurosci.* **29**, 298–306 (2006).
- Sillito, A. M., Cudeiro, J. & Jones, H. E. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends Neurosci.* **29**, 307–316 (2006).
- Bolz, J. & Gilbert, C. D. Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature* **320**, 362–365 (1986).
- Grieve, K. L. & Sillito, A. M. A re-appraisal of the role of layer VI of the visual cortex in the generation of cortical end inhibition. *Exp. Brain Res.* **87**, 521–529 (1991).
- Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
- Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100**, 13940–13945 (2003).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154 (1962).
- Chow, B. Y. *et al.* High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
- Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165 (2010).
- Rafols, J. A. & Valverde, F. The structure of the dorsal lateral geniculate nucleus in the mouse. A Golgi and electron microscopic study. *J. Comp. Neurol.* **150**, 303–331 (1973).
- Salinas, E. & Thier, P. Gain modulation: a major computational principle of the central nervous system. *Neuron* **27**, 15–21 (2000).
- Brothie, P. R., Andersen, R. A., Snyder, L. H. & Goodman, S. J. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* **375**, 232–235 (1995).
- Treue, S. & Martinez Trujillo, J. C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579 (1999).
- McAdams, C. J. & Maunsell, J. H. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
- Silver, R. A., Lubke, J., Sakmann, B. & Feldmeyer, D. High-probability unquantal transmission at excitatory synapses in barrel cortex. *Science* **302**, 1981–1984 (2003).
- Adesnik, H. & Scanziani, M. Lateral competition for cortical space by layer-specific horizontal circuits. *Nature* **464**, 1155–1160 (2010).
- Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5**, 793–807 (2004).
- Ascoli, G. A. *et al.* Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Rev. Neurosci.* **9**, 557–568 (2008).
- Chance, F. S., Abbott, L. F. & Reyes, A. D. Gain modulation from background synaptic input. *Neuron* **35**, 773–782 (2002).
- Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
- Murphy, B. K. & Miller, K. D. Multiplicative gain changes are induced by excitation or inhibition alone. *J. Neurosci.* **23**, 10040–10051 (2003).
- Andolina, I. M., Jones, H. E., Wang, W. & Sillito, A. M. Corticothalamic feedback enhances stimulus response precision in the visual system. *Proc. Natl Acad. Sci. USA* **104**, 1685–1690 (2007).
- Wang, W., Jones, H. E., Andolina, I. M., Salt, T. E. & Sillito, A. M. Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neurosci.* **9**, 1330–1336 (2006).
- Wörgötter, F., Nelle, E., Li, B. & Funke, K. The influence of corticofugal feedback on the temporal structure of visual responses of cat thalamic relay cells. *J. Physiol.* **509**, 797–815 (1998).
- McClurkin, J. W. & Marrocco, R. T. Visual cortical input alters spatial tuning in monkey lateral geniculate nucleus cells. *J. Physiol.* **348**, 135–152 (1984).
- Murphy, P. C., Duckett, S. G. & Sillito, A. M. Feedback connections to the lateral geniculate nucleus and cortical response properties. *Science* **286**, 1552–1554 (1999).
- Casagrande, V. A. & Kaas, J. H. *The Afferent, Intrinsic and Efferent Connections of Primary Visual Cortex in Primates* (eds Peters, A. & Rockland, P.) (Plenum, 1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to M. Carandini, J. Isaacson and the members of the Scanziani and Isaacson laboratories for helpful discussions of this project, to J. Isaacson, R. Malinow and T. Komiyama for providing feedback on the manuscript, to P. Abelkop for histological help and neonatal viral injections, to J. Evora for mouse colony support and genotyping, to B. Atallah for sharing the technique for silencing the cortex by photostimulation of parvalbumin neurons and for help with the *in vivo* recording setup and to W. Bruns for help coding analysis software. We thank the UCSD Neuroscience Microscopy Facility (P30 NS047101) for the use of their imaging equipment. S.R.O. and H.A. were supported by postdoctoral fellowships from the Helen Hay Whitney Foundation. D.S.B. was supported by a UCSD Neurobiology Training Grant (NINDS: 5T32NS007220-28). M.S. is an investigator of the Howard Hughes Medical Institute. This work was also supported National Institutes of Health grant RO1 NS069010 and by the Gatzsby Charitable Foundation.

Author Contributions H.A. performed the initial physiological characterization of the NTSR1-Cre expression system with optogenetic tools. H.A. also developed the *in vivo* awake recording preparation on the treadmill. S.R.O. performed all *in vivo* recordings. D.S.B. performed all *in vitro* recordings and anatomical reconstructions. S.R.O. and M.S. designed the study. M.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.S. (massimo@biomail.ucsd.edu) or S.R.O. (srolsen@ucsd.edu).

METHODS

All procedures were conducted in accordance with the National Institutes of Health guidelines and with the approval of the Committee on Animal Care at the University of California, San Diego.

Animals. We used the following mouse lines: NTSR1-Cre (strain B6.FVB(Cg)-Tg(Ntsr1-cre)GN220Gsat/Mmcd, stock number 030648-UCD), which was generated by the GENSAT project¹⁹ and acquired from the Mutant Mouse Regional Resource Centers; tdTomato reporter (Jax number 007908); GAD67-GFP (Aneo); and PV-Cre (Silvia Arber).

Viral injections. Adeno-associated viruses (AAVs) for ChR2 and archaerhodopsin were acquired from the University of Pennsylvania Viral Vector Core: AAV2/1.CAGGS.flex.ChR2.tdTomato.SV40 (Addgene 18917) and AAV2/9.flex.CBA.Archaerhodopsin-GFP.W.SV40 (Addgene 22222). An AAV virus (AAV2/9) for NpHR3.0 was produced at the Salk Viral Vector Core. The NpHR3.0 plasmid (pAAV-Efla-DIO-eNpHR 3.0-EYFP) was provided by K. Diesseroth.

Viruses were loaded in a beveled sharp micropipette mounted on a Nanoject II (Drumond) or a micropump injector (UMP-3 WPI) attached to a micromanipulator. ChR2 virus was injected into newborn pups (between postnatal days 0 and 2) that were anaesthetized on ice and secured into a moulded platform. Three 20-nl boli of virus was injected at each of three medial-lateral locations in V1 and two depths (500 μm and 650 μm) within V1.

Archaerhodopsin was injected in combination with NpHR3.0 in juvenile (1–2-month-old) mice anaesthetized with 2.5% isoflurane and placed into a stereotaxic frame (Knopf). The exposed skull overlying V1 was thinned in three locations with a dental drill (Foredom) with a 300- μm bur (Gesswein), and a hole was made with a (25-gauge) needle at each location to permit insertion of the injection pipette. A volume of 150 nl of virus was injected at a rate of 20 nl min⁻¹ at each of the three locations and at two depths (900 μm and 700 μm). The scalp was then sutured and the mouse injected subcutaneously with 0.1 mg kg⁻¹ buprenorphine. *In vivo* recordings were made 1–2 months after viral injection.

Slice preparation. Mice were anaesthetized with ketamine and xylazine (100 mg kg⁻¹ and 10 mg kg⁻¹, respectively), perfused transcardially with cold sucrose solution (in mM: NaCl, 83; KCl, 2.5; MgSO₄, 3.3; NaH₂PO₄, 1; NaHCO₃, 26.2; D-glucose, 22; sucrose, 72; and CaCl₂, 0.5, bubbled with 95% O₂ and 5% CO₂) and decapitated, and the visual cortex was cut into 300–400- μm coronal sections in cold sucrose solution. Thalamic slices were cut 45° off the coronal plane to maintain connections between NRT and dLGN. Slices were incubated in sucrose solution in a submerged chamber at 34 °C for 30 min and then at room temperature (21 °C) until used for recordings.

***In vitro* recordings.** Whole-cell recordings were done at 32 °C in artificial cerebrospinal fluid (in mM: NaCl, 119; KCl, 2.5; NaH₂PO₄, 1.3; NaHCO₃, 26; D-glucose, 20; MgCl₂, 1.3; CaCl₂, 2.5; and mOsm, 305, bubbled with 95% O₂ and 5% CO₂). Excitatory and inhibitory synaptic currents were recorded using a caesium-based internal solution (in mM: CsMeSO₄, 115; NaCl, 4; HEPES, 10; Na₃GTP, 0.3; MgATP, 4; EGTA, 0.3; QX-314-Cl, 2.5; BAPTA(5Cs), 10; adjusted to pH 7.4 with CsOH; mOsm 295; 3–5 MOhm pipette resistance). Voltage-clamp recordings were not considered if the series resistance exceeded 20 MOhm or varied by more than 10%. Typically, 2–4 neurons were recorded from simultaneously. Cell-attached recordings and biocytin fills were carried out with a potassium-based internal solution (in mM: K-gluconate, 150; MgCl₂, 1.5; HEPES, 5; EGTA, 1.1; phosphocreatine, 10; adjusted to pH 7.4 with KOH; mOsm 295). Perforated-patch recordings were carried out using potassium-based internal and 10 $\mu\text{g ml}^{-1}$ Gramicidin D (Sigma G5002). Tight seals were held until sufficient access allowed injection of current and resolution of action potentials (typically 10–20 min). Ruptures of the perforated patch were apparent by a rapid drop in series resistance at which point the recordings were discontinued. Photostimulation of L6 *in vitro* consisted of either single 2-ms pulses or a 40-Hz train of 2-ms pulses, or of 1-s ramps of light of increasing intensity as previously described³². Data were recorded with Multiclamp 700B amplifiers (Axon instruments) filtered at 2 kHz and digitized with a Digidata1440A (Axon instruments) at 10 kHz. Recordings were analysed using custom-made routines in Igor Pro (Wavemetrics). Charges represent the time integral of the synaptic current recorded during the first second of photostimulation. The stage was moved using a custom made plugin for ImageJ(NIH) to interface with ESP300 (Newport) via SerialPort (SerialIO). Drugs used were NBQX (Tocris 1044) and CPP (Ascent Asc-159).

***In vivo* recordings in anaesthetized mice.** Recordings were performed similarly to those previously described²². Animals were anaesthetized with 5 mg kg⁻¹ of chlorporthixene (intraperitoneal) and then (5–10 min later) with 1.2 g kg⁻¹ urethane (intraperitoneal). During surgery, animals were given 0.5–1.0% isoflurane. Animals were placed onto a custom platform and their temperature was maintained at 37 °C using a feedback-controlled heating pad (FHC). Whiskers and eyelashes that were contralateral to the recording side were trimmed and eyes covered with a

thin, uniform layer of silicone oil to prevent drying. Protein expression was verified by transcranial epifluorescence of the exposed and PBS-moistened skull using a Leica MZ10F microscope. Only animals showing expression over the entire extent of V1 were used for subsequent experiments. The entire dried skull was covered with black dental cement (Ortho-Jet powder (Lang Dental) mixed with black iron oxide) but for the previously outlined boundaries of V1 (~1.5–3.5 mm lateral to midline and –0.5 to 2.5 mm anterior to lambda suture). A head-plate with a hole of ~2 mm in diameter was mounted over V1 and a small region of skull (~300 × 750 μm) was thinned using a dental drill. Next, we used sharpened fine forceps (Dumont number 55) to make a craniotomy just sufficiently large for inserting the probe. A drop of PBS placed in the well at the centre of the head-plate kept the exposed skull and craniotomy moist. A multichannel silicon probe mounted on a micromanipulator (Luigs–Neumann) was slowly advanced into the brain to a depth of 800–1000 μm for linear probes and 200–700 μm for tetrode probes (see later), and recordings were started 20 min or more after inserting the probe.

For dLGN recordings we made a circular craniotomy (~1.5 mm in diameter) 2.6 mm posterior and 2 mm lateral to the bregma suture. Robust visual responses and bursting activity that was characteristic of dLGN relay neurons were encountered at a depth between 2,400 and 3,100 μm^{44} (Supplementary Fig. 7). For dual recording experiments (Fig. 3g, h and Fig. 5), we used a larger head-plate so that a craniotomy could be made both over the dLGN and V1.

Recordings were made with NeuroNexus 16-channel linear (a1x16-3mm-50-177) or tetrode (a2x2-tet-3mm-150-121) silicon probes. For recordings across cortical depth and in dLGN we used the linear configuration. The tetrode configuration was used to isolate a subset of cells in Fig. 2. Signals were amplified ×1000, band-pass filtered between 0.3 Hz and 5 kHz using an AM Systems 3500 amplifier and acquired at 32 kHz using a NIDAQ board (PCIe-6239) controlled with custom-written software in Matlab (Mathworks). For dual recording experiments we used two separate data-acquisition setups (amplifier, NIDAQ board and computer). Raw data were stored on a computer hard drive for offline analysis.

At the end of the recording session, animals were killed by administering 4% isoflurane and the brain was quickly removed and fixed in 4% paraformaldehyde for histological analysis.

***In vivo* awake recordings.** 1–2 weeks before recording, mice were implanted with a head-plate for head fixation. Mice were anaesthetized with 2.5% isoflurane, the scalp was removed and a head-plate was fixed over V1 with black dental cement. The skull directly overlying V1 was covered with Kwik-Cast (WPI). Animals were injected subcutaneously with 0.1 mg kg⁻¹ buprenorphine and allowed to recover in their home cage for at least 1 week before recording.

Several days before recording, mice were familiarized to head fixation within the recording setup. They were briefly anaesthetized with isoflurane and the head-plate was clamped to a metal post, but otherwise the mice were unrestrained and allowed to run in this position on a plastic circular treadmill or track (Fast-Trac from Bio-Serv; see Supplementary Fig. 6). The same circular track was present in the cages of the mice, where they were familiarized with its use. Mice grew accustomed to head fixation over the course of 1–3 15-min sessions and ran naturally on the track, occasionally stopping to rest or groom.

On the day of recording, mice were anaesthetized with 1.5–2% isoflurane, a small craniotomy was made over V1, a drop of PBS was placed in the well of a head-plate that was clamped to a metal post, and the multichannel probe inserted into the craniotomy. After removal of isoflurane the mice regained consciousness and typically began running. Recordings did not start before 30 min after the end of anaesthesia. Awake recording sessions lasted between 1 and 2 h. Mice typically spent ~60–80% of their time running, and the rest of the time was spent resting or grooming. Data were not separated according to behaviour. Every 30–60 min mice were given a few drops of a 5% glucose solution through a disposable pipette. For two mice we performed 2–3 recording sessions, which were made at least a day apart. Between sessions the craniotomy was covered with Kwik-Cast. A new craniotomy was made for each session.

Visual stimulation. Visual stimuli were generated in Matlab using the Psychophysics Toolbox⁴⁵ and were displayed on a gamma-corrected LCD monitor (Dell 52 × 32.5 cm, 60-Hz refresh rate, mean luminance 50 cd m⁻²) positioned 25 cm from the contralateral eye. The monitor was positioned for each experiment so that the multi-unit receptive field was located approximately in the centre of the screen (the multi-unit receptive field was determined by moving a localized drifting grating patch (~10°) around the screen). During the recording session full-field sinusoidal drifting gratings were used. All stimuli had a temporal frequency of 2 Hz and a spatial frequency of 0.04 cycles per degree. Gratings were randomly presented at 8–12 equally spaced directions, except for the experiments in Fig. 5 in which we used only two orthogonal grating directions (0° and 90°). The contrast of the stimulus was 100%, except for Fig. 5 in which we used six contrast levels (2, 4.4,

9.6, 21, 46 and 100%). A grey screen trial was interleaved with the drifting gratings. The duration of the visual stimulus was 1.5 s and the inter-trial interval was 3–6 s. **In vivo photostimulation.** To photo-stimulate ChR2 we used a blue (470-nm) fibre-coupled LED (1 mm diameter, Doric Lenses) placed ~5–10 mm away from the skull. Light from the LED spanned the entire area of V1. An opaque shield of black aluminium foil (Thor Labs) prevented LED light from reaching the contralateral eye. The LED was driven by the analogue output from the NIDAQ board. The blue LED was presented at five intensities (approximately 3, 5, 7, 10.5 and 20 mW measured at the tip of the fibre), but for a minority of experiments we presented only the highest LED intensity. Trials were alternated between visual stimulus only and visual stimulus plus LED. The strongest LED intensity also generated oscillations at gamma frequency, consistent with previous observations³² (Supplementary Fig. 2). The preferred-orientation of photostimulated L6 cells remained unchanged but their tuning curves became broader (Supplementary Fig. 2).

To photostimulate archaerhodopsin and NpHR3.0 we used an amber (590-nm) fibre-coupled LED (1 mm in diameter, Doric Lenses) placed ~0.5 mm from the skull. Because photosuppression of L6 produced a transient decrease in spontaneous multi-unit activity in L2–5 at the onset of LED illumination (see Supplementary Fig. 6) we turned on the amber LED 1.4 s before the visual stimulus began. Experiments were performed at the highest LED intensity (~20 mW measured at the tip of the fibre). As long as the suppression was not complete, the preferred orientation of photosuppressed L6 cells remained unchanged (Supplementary Fig. 6). **In vivo data analysis.** All *in vivo* data analysis was performed with custom software written in Matlab.

Multi-unit spiking activity was defined as all events (spikes) exceeding a threshold of 4 s.d. above the noise of the high-pass filtered (500-Hz) signal. Spikes were assigned a depth corresponding to the depth of the channel they were recorded from. Spikes that were recorded simultaneously on multiple channels were considered as a single event and attributed to the channel in which they showed the largest amplitude. We determined the depth of each channel by considering the depth and the angle of the probe relative to the vertical axis of cortex. We assigned spikes to different layers according to the following depths (in μm): L2/3, 100–350; L4, 350–450; L5, 450–650; L6, >650. PSTHs were composed of 50-ms bins. PSTHs of individual experiments were normalized to the first 500 ms of the visual stimulus (for ChR2 experiments) or to the entire visual stimulus (for archaerhodopsin and NpHR3.0 experiments) to generate average PSTHs. PSTHs for kinetic analysis (Fig. 3h) were composed of 3-ms bins and report the normalized difference in firing rates between control (average firing over a 50-ms window prior to LED onset) and during LED illumination (average firing rate over a 100-ms window, 50 ms after LED onset). For each experiment the onset of suppression was determined as the time point at which the normalized response fell below 2 s.d. of the baseline.

The contrast response functions in dLGN and V1 report the normalized, baseline-subtracted firing rates and were fitted with a hyperbolic ratio function:

$$r = r_{\max} \frac{c^n}{c^n + c_{50}^n}$$

where r is the response, c is the contrast of the visual stimulus, r_{\max} is a fitted constant representing the response saturation level, n is fitting exponent that affects the shape of the curve and c_{50} is the semi-saturation constant. The transfer function between the dLGN and V1 was fitted with a hyperbolic ratio function:

$$r_{V1} = r_{V1, \max} \frac{r_{\text{dLGN}}^n}{r_{\text{dLGN}}^n + r_{\text{dLGN}, 50}^n}$$

where r_{V1} is the V1 response, $r_{V1, \max}$ is a constant representing the V1 saturation level, r_{dLGN} is the dLGN response, n is a fitting exponent and $r_{\text{dLGN}, 50}$ is the semi-saturation constant. The 'corticothalamic component' (CT) was defined as the fraction of the total V1 suppression accounted for by this predicted response. The 'intracortical component' was then defined as $1 - \text{CT}$ component. We performed this analysis for five LED levels and averaged across experiments to produce the plot in Fig. 5f.

We isolated single units using spike-sorting software provided by D. N. Hill, S. B. Mehta, and D. Kleinfeld⁴⁶. For both the linear and tetrode probes we analysed waveforms extracted from groups of four adjacent electrode sites. We high-pass filtered the raw signal at 500 Hz and then detected spiking events exceeding 4–5 s.d. of the noise. Spike waveforms were clustered using a k -means algorithm. After initial automated clustering, we used a graphical user interface to manually merge

and split clusters. Unit isolation quality was assessed by considering refractory period violations and Fisher linear discriminant analysis. In agreement with previous studies we could classify waveforms as regular-spiking or fast-spiking putative inhibitory neurons. In our data set there was a clear bimodal distribution of trough-to-peak times (a threshold of 0.4 ms was used to divide fast-spiking from regular-spiking units). All units were assigned a depth according to the channel that they were detected on, and units were assigned to layers based on the depth divisions given above for the multi-unit activity.

For each unit we computed the visual response as the mean spike-rate occurring over the time window in which both the LED and visual stimulus were present. Thus, for the L6 photostimulation experiments this typically corresponded to a 500-ms window placed in the centre of the visual response, and for the L6 photosuppression experiments this window encompassed the entire 1.5-s visual stimulus. For all analysis except the orientation tuning analysis in Fig. 2, we averaged responses over all stimulus conditions. Following recent studies^{47,48} of orientation tuning we computed an OSI as:

$$\text{OSI} = \frac{\sqrt{(\sum r_k \sin(2\theta_k))^2 + (\sum r_k \cos(2\theta_k))^2}}{\sum r_k}$$

where r_k is the response to the k th direction given by θ_k . We determined an OSI for each unit with and without photostimulation or suppression of L6. We established the preferred orientation and tuning width by first fitting the average responses of each unit with a sum of two Gaussians:

$$r = r_0 + r_p e^{-(\theta - \theta_p)^2 / (2\sigma^2)} + r_{p+180} e^{-(\theta - \theta_p - 180)^2 / (2\sigma^2)}$$

where r_0 is a constant offset, r_p is the response at the preferred orientation, r_{p+180} is a response 180° away from the preferred direction, θ is the stimulus direction, θ_p is the preferred orientation and σ is the tuning width. The two Gaussians were forced to peak 180° apart and to have the same width but could have different amplitudes. Control and photostimulation or photosuppression conditions were fit separately. To generate the average population tuning curve we first circularly shifted the stimulus direction of each unit so that the maximal response occurred at 0° . We then normalized the responses to this peak response and averaged all normalized tuning curves together. We fit the control population average tuning curve with a sum of two Gaussians. The curve for the photostimulation or photosuppression population average was produced by scaling the control curve by the slope (gain factor) of the linear fit shown in Fig. 2e, j.

All error bars are presented as mean \pm s.e.m. unless otherwise noted. We used paired t -tests to assess statistical significance unless otherwise noted.

Histology. Triple transgenic mice (Ntsr1-Cre, floxed-tdTomato and Gad67-GFP) were anaesthetized with ketamine and xylazine (100 mg kg⁻¹ and 10 mg kg⁻¹, respectively) and perfused with cold sucrose (see above) and then perfluoroalkoxy (4% in PBS). After 24 h incubation in perfluoroalkoxy, slices were cut into 50- μm sections and immunostained as described previously⁴⁹. Antibodies that were used were mouse anti-NeuN (1:400; Millipore MAB377), chicken anti-GFP (1:1000; Aves Labs GFP-1020), goat anti-chicken AF488 (1:1,000; Invitrogen A11039) and goat anti-mouse AF633 (1:1,000; Invitrogen A21050). Slices were mounted in Vectashield with Dapi (Vector Labs, H1500). Images were single confocal sections taken on an Olympus FV1000. Layer borders were identified by changes in cell density. Cell counts were carried out using standard stereological techniques. Biocytin fills and neural reconstructions were done as previously described⁵⁰.

44. Grubb, M. S. & Thompson, I. D. Quantitative characterization of visual response properties in the mouse dorsal lateral geniculate nucleus. *J. Neurophysiol.* **90**, 3594–3607 (2003).
45. Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
46. Fee, M. S., Mitra, P. P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J. Neurosci. Methods* **69**, 175–188 (1996).
47. Kerlin, A. M., Andermann, M. L., Berezovskii, V. K. & Reid, R. C. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron* **67**, 858–871 (2010).
48. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation selectivity in macaque V1: diversity and laminar dependence. *J. Neurosci.* **22**, 5639–5651 (2002).
49. Bortone, D. & Polleux, F. KCC2 expression promotes the termination of cortical interneuron migration in a voltage-sensitive calcium-dependent manner. *Neuron* **62**, 53–71 (2009).
50. Bagnall, M. W., Hull, C., Bushong, E. A., Ellisman, M. H. & Scanziani, M. Multiple clusters of release sites formed by individual thalamic afferents onto cortical interneurons ensure reliable transmission. *Neuron* **71**, 180–194 (2011).

Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes

Jennifer F. Hughes¹, Helen Skaletsky¹, Laura G. Brown¹, Tatyana Pyntikova¹, Tina Graves², Robert S. Fulton², Shannon Dugan³, Yan Ding³, Christian J. Buhay³, Colin Kremitzki², Qiaoyan Wang³, Hua Shen³, Michael Holder³, Donna Villasana³, Lynne V. Nazareth³, Andrew Cree³, Laura Courtney², Joelle Veizer², Holland Kotkiewicz², Ting-Jan Cho¹, Natalia Koutseva¹, Steve Rozen¹, Donna M. Muzny³, Wesley C. Warren², Richard A. Gibbs³, Richard K. Wilson² & David C. Page¹

The human X and Y chromosomes evolved from an ordinary pair of autosomes during the past 200–300 million years^{1–3}. The human MSY (male-specific region of Y chromosome) retains only three percent of the ancestral autosomes' genes owing to genetic decay^{4,5}. This evolutionary decay was driven by a series of five 'stratification' events. Each event suppressed X–Y crossing over within a chromosome segment or 'stratum', incorporated that segment into the MSY and subjected its genes to the erosive forces that attend the absence of crossing over^{2,6}. The last of these events occurred 30 million years ago, 5 million years before the human and Old World monkey lineages diverged. Although speculation abounds regarding ongoing decay and looming extinction of the human Y chromosome^{7–10}, remarkably little is known about how many MSY genes were lost in the human lineage in the 25 million years that have followed its separation from the Old World monkey lineage. To investigate this question, we sequenced the MSY of the rhesus macaque, an Old World monkey, and compared it to the human MSY. We discovered that during the last 25 million years MSY gene loss in the human lineage was limited to the youngest stratum (stratum 5), which comprises three percent of the human MSY. In the older strata, which collectively comprise the bulk of the human MSY, gene loss evidently ceased more than 25 million years ago. Likewise, the rhesus MSY has not lost any older genes (from strata 1–4) during the past 25 million years, despite its major structural differences to the human MSY. The rhesus MSY is simpler, with few amplified gene families or palindromes that might enable intrachromosomal recombination and repair. We present an empirical reconstruction of human MSY evolution in which each stratum transitioned from rapid, exponential loss of ancestral genes to strict conservation through purifying selection.

The human Y chromosome no longer engages in crossing over with its once-identical partner, the X chromosome, except in its pseudoautosomal regions. During evolution, X–Y crossing over was suppressed in five different chromosomal regions at five different times, each probably resulting from an inversion in the Y chromosome^{2,3}. Each of these regions of the Y chromosome then began its own individual course of degeneration, experiencing deletions and gene loss. Comparison of the present-day X and Y chromosomes enables identification of these five evolutionary 'strata' in the MSY (and X chromosome); their distinctive degrees of X–Y differentiation indicate their evolutionary ages^{2,3}. The oldest stratum (stratum 1) dates back over 240 million years (Myr)² and is the most highly differentiated, and the youngest stratum (stratum 5) originated only 30 Myr ago and displays the highest X–Y nucleotide sequence similarity within the MSY³. The five strata and their respective decay processes, over tens to hundreds of millions of years of mammalian evolution, offer replicate experiments of nature from which to reconstruct the trajectories and kinetics of gene loss in the MSY.

Only the human and chimpanzee MSYs had been sequenced before the present study, and they are separated by just 6 Myr of evolution. We decided to examine the MSY of a much more distant relative, the rhesus macaque (*Macaca mulatta*), to enable us to reconstruct gene loss and conservation in the MSY during the past 25 Myr. We sequenced the rhesus MSY using bacterial artificial chromosome (BAC) clones and the SHIMS (single-haplotype iterative mapping and sequencing) strategy that has previously been used in the human and chimpanzee MSYs^{4,11–13} as well as in the chicken Z chromosome⁵. The resulting sequence is comprised of 11.0 megabases (Mb), is complete aside from three small gaps and has an error rate of about one nucleotide per Mb. We ordered and oriented the finished sequence contigs by fluorescence *in situ* hybridization and radiation hybrid mapping (Supplementary Figs 1–6, Supplementary Table 1, Supplementary Files 1, 2 and Supplementary Note 1).

We then compared the structure of the rhesus Y chromosome to that of the human and chimpanzee (Fig. 1). The rhesus Y chromosome has virtually no heterochromatin apart from the centromere, and the euchromatic segment of the MSY is notably smaller compared to that of the human and chimpanzee (Fig. 1). The single pseudoautosomal region (PAR) in rhesus corresponds to the short-arm PAR in human and to the single PAR in chimpanzee. The precise boundary between PAR and MSY is identical in the three species (Supplementary Fig. 7), confirming that stratification in all three lineages concluded before the divergence of apes from Old World monkeys.

The euchromatic portions of the rhesus, human and chimpanzee MSYs are comprised primarily of two distinct sequence classes: X-degenerate and ampliconic. The X-degenerate regions, relics of shared X–Y ancestry, are dotted with single-copy homologues of X-linked genes. The X-degenerate regions are relatively well conserved among the rhesus, human and chimpanzee MSYs, with large blocks of homology that are readily identifiable (Supplementary Figs 8 and 9). Indeed, the X-degenerate regions are the only portions of the rhesus and human MSYs whose sequences can be aligned over distances of greater than 50 kb. We found rhesus–human nucleotide divergence there to be 9.4% (Supplementary File 3). This is markedly higher than the 6.5% divergence that is observed when the rhesus and human female genomes are compared¹⁴. The difference probably reflects the restriction of the MSY to the male germ line, where base-pair substitutions are more frequent than in the female germ line¹⁵. From these data, we calculate the male-to-female mutation rate ratio (α_m) to be 2.78 (95% confidence interval 2.74–2.81), in agreement with previous but less precise estimates^{14,16}. The X-degenerate sequences in rhesus, human and chimpanzee are not entirely colinear, as large-scale rearrangements have occurred in each lineage (Supplementary Figs 8–10).

For all three species, the MSY's ampliconic regions are composed of long, nearly identical repeat units that are arrayed in either direct or

¹Howard Hughes Medical Institute, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²The Genome Institute, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, Missouri 63108, USA. ³Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA.

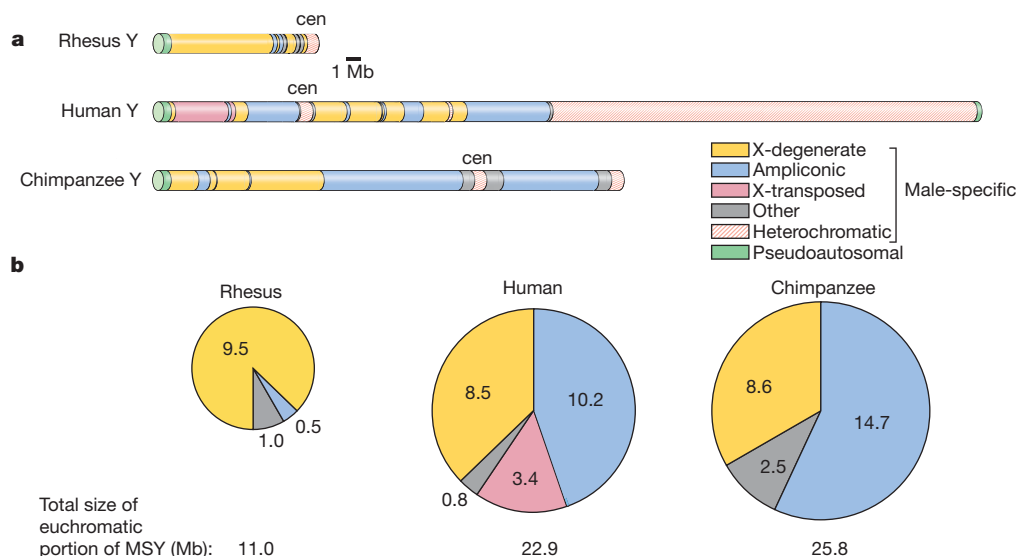


Figure 1 | Comparison of rhesus, human and chimpanzee Y chromosomes. **a**, Schematic representations of rhesus, human and chimpanzee Y chromosomes, to scale. Other, single-copy, male-specific sequences that are

inverted orientation and undergo frequent gene conversion—a process that is thought to slow or prevent the decay of genes that reside there^{4,17}. Ampliconic genes display testis-specific expression patterns, consistent with their having critical roles in spermatogenesis^{4,11,18}. Only 0.5 Mb of the rhesus MSY euchromatin is ampliconic, compared to 10.2 Mb and 14.7 Mb in human and chimpanzee, respectively (Fig. 1, and Supplementary Figs 11 and 12). In human and chimpanzee, the ampliconic regions of the MSY feature large palindromes, each composed of two inverted repeats (arms) separated by a short spacer. The human and chimpanzee MSYs have 8 and 19 palindromes that span 5.5 Mb and 7.5 Mb, respectively^{4,13}. By contrast, the rhesus MSY has only three palindromes and these collectively span 437 kb (Supplementary Table 2 and Supplementary Fig. 13). Two of the rhesus MSY palindromes are orthologues of human MSY palindromes, demonstrating that these structures have been maintained for at least 25 Myr (Supplementary Fig. 13).

We identified protein-coding genes in the rhesus MSY using three complementary approaches. First, we electronically searched the rhesus MSY for homologues of all known human and chimpanzee MSY genes and pseudogenes. Second, we searched for homologues of all known human X-linked genes, to identify any X–Y shared genes that had been lost in both the human and chimpanzee MSY but retained in the rhesus MSY. Third, we searched for additional rhesus-specific MSY genes using electronic prediction tools and high-throughput sequencing of rhesus testis complementary DNA (245 Mb in total). We validated each putative gene by verifying transcriptional activity (Supplementary Fig. 14) and, where applicable, by comparing its predicted open reading frame to that of its human orthologue (Supplementary Table 3).

We then compared the catalogues of MSY genes in rhesus, human⁴ and chimpanzee¹³ to infer gene loss and conservation during the past 25 Myr. To root this analysis in a deep evolutionary context, we first reconstructed which of the modern rhesus MSY genes were present on the common autosomal ancestor of X and Y (Fig. 2, Supplementary Table 4 and Supplementary Note 2). Most ‘ancestral’ MSY genes would be expected to have a homologue both on the human X chromosome and on the chicken autosomes (chromosomes 1 and 4) that share common ancestry with mammalian X and Y chromosomes^{3,5}. Indeed, 33 genes and pseudogenes in the rhesus, human or chimpanzee MSY have their most closely related human homologues on the X chromosome (Fig. 2), and 29 of these also have homologues within syntenic regions of chicken chromosome 1 or 4. Analyses of a more distant outgroup,

neither X-degenerate nor X-transposed. **b**, Sizes (in Mb) of euchromatic sequence classes in MSYs. cen, centromere.

Xenopus tropicalis, revealed that two of the four rhesus MSY genes lacking homologues on chicken chromosome 1 and 4 (*TSPY* and *AMELY*) are X–Y ancestral; they were lost in the chicken lineage after divergence from mammals (Supplementary Note 2). A few human MSY genes with X homologues are recent additions to the MSY rather than remnants of the ancestral autosome pair; *PCDH11Y* and *TGIF2LY* are located in the human-specific X-transposed region⁴, and the X-linked homologue of *VCY* is found only in simian primates¹⁹. We found a total of 30 ancestral MSY genes and pseudogenes in rhesus, human or chimpanzee (Fig. 2).

Within strata 1–4, which collectively comprise the bulk of the human MSY, the rhesus and human MSYs possess precisely the same 18 ancestral genes (Fig. 2). This notable and unanticipated identity leads us to conclude that, 25 Myr ago, in the last common ancestor of rhesus and human, MSY strata 1–4 also carried these 18 ancestral genes (Table 1 and Supplementary Table 5), and that no loss of ancestral genes occurred subsequently in either lineage (Supplementary Note 3). We note that, within strata 3 and 4, the rhesus and human MSYs carry a total of six ancestral pseudogenes that seem to have lost their function more than 25 Myr ago (Supplementary Fig. 15).

The evolutionary stability of ancestral genes in strata 1–4 could be explained by purifying selection, which, in the absence of sexual recombination, would have preserved critical ancestral genes for tens or even hundreds of millions of years. We demonstrated previously that purifying selection preserved MSY genes during the past 100,000 years of human population expansion and migration²⁰. Comparing human and rhesus, we find that most ancestral genes display a ratio of nonsynonymous substitution rate to synonymous substitution rate that is significantly less than one (Supplementary Note 4, Supplementary Table 3 and Supplementary Fig. 16), demonstrating purifying selection during the past 25 Myr.

The pattern of gene loss and conservation in stratum 5, formed only 5 Myr before the rhesus and human lineages split, is remarkably different from the pattern in the four older strata. Within the past 30 Myr, four ancestral genes have been inactivated or deleted from stratum 5 of the MSY in both rhesus and human (Fig. 2 and Supplementary Note 5). A fifth ancestral gene, *MXRA5Y*, remains active in rhesus (Supplementary Fig. 11) but was inactivated by an intragenic deletion in the human lineage (Supplementary Fig. 17). Apart from *MXRA5Y*, all differences in MSY gene content between rhesus and human involve genes that were added to the human MSY subsequent to the ape–Old World monkey split (Fig. 2 and Supplementary Table 5).

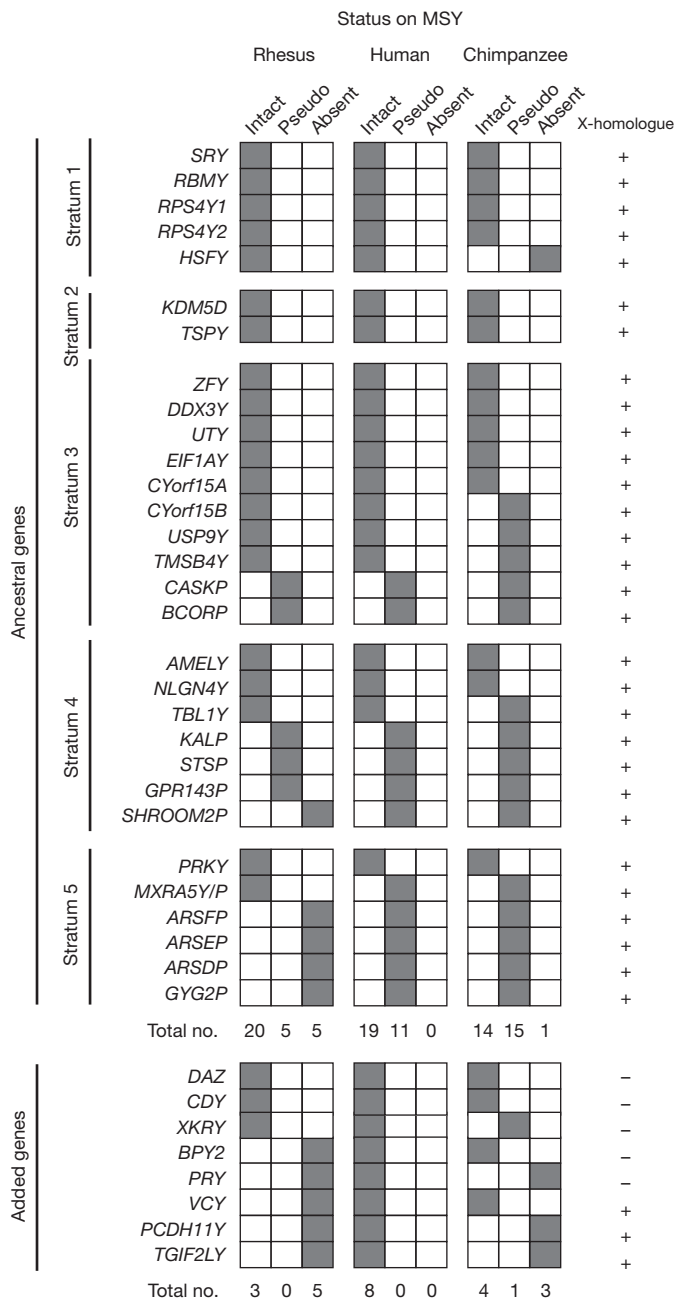


Figure 2 | Inventories of genes, both ancestral and added, in rhesus, human and chimpanzee MSYs. Ancestral genes grouped by stratum (1–5). In rhesus, human and chimpanzee, current status of each MSY gene is indicated by shading in one of three columns: present and intact, inactivated pseudogene, or absent or deleted. Total numbers of intact genes, pseudogenes (pseudo), and absent genes—both ancestral and added—are tallied for each species. For each MSY gene, whether the most closely related human homologue is located on the X chromosome is shown (right).

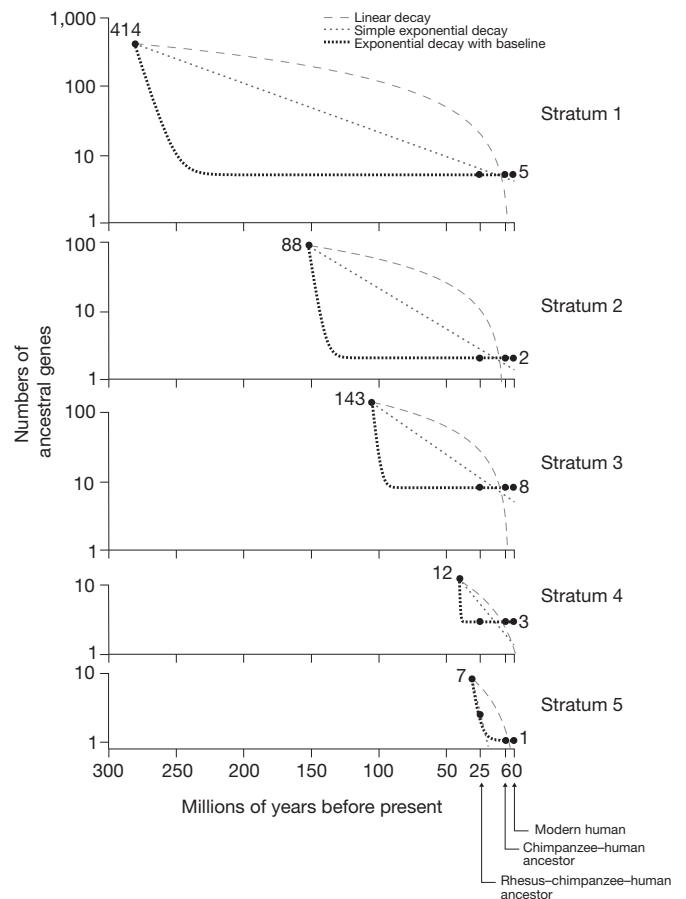


Figure 3 | Kinetics of ancestral gene loss during evolution of five human MSY strata. Gene numbers are plotted on a log scale on the y axis, and time (in Myr before present) is plotted on the x axis. Filled circles show inferred or observed gene numbers in (from left to right) X–Y ancestral chromosome (at time of stratum formation), rhesus–chimpanzee–human ancestral MSY (25 Myr ago), chimpanzee–human ancestral MSY (6 Myr ago), and modern human MSY. Dotted and dashed lines represent best-fit curves to data points using each of three decay models as indicated.

Returning to strata 1–4, we note that five ancestral genes have been inactivated or lost from the chimpanzee MSY during the past 6 Myr^{12,13}, in sharp contrast to the strict conservation of ancestral gene content in rhesus and human (Fig. 2). We previously proposed that in the chimpanzee lineage promiscuous mating behaviour²¹, sperm competition and intense sexual selection that focused on the MSY drove rapid evolution and amplification of MSY sequences that are associated with spermatogenesis^{12,13}. Furthermore, we speculated that in the chimpanzee lineage inactivated alleles of some ancestral genes became fixed in the population through ‘genetic hitchhiking’; casualties of positive but indiscriminate selection operating in the absence of sexual recombination in the MSY^{12,13,22}. Among primate species, chimpanzees have a high testis-weight to body-weight ratio, a useful indicator of the degree of sperm competition^{23,24}. Although the rhesus

Table 1 | Stratification of X–Y ancestral gene loss in primate MSYs

	Age of stratum (millions of years) (from refs 2, 3)	Number of ancestral genes on human X chromosome*	Number of ancestral genes on MSY			
			Last common ancestor†	Rhesus	Human	Chimpanzee
Stratum 1	240–320	414	5	5	5	4
Stratum 2	130–170	88	2	2	2	2
Stratum 3	80–130	143	8	8	8	5
Stratum 4	38–44	12	3	3	3	2
Stratum 5	29–32	7	2–7	2	1	1

* Gene numbers from ref. 5, Supplementary Table 4 and Supplementary Note 2.

† Gene counts in MSY of a hypothetical rhesus–human–chimpanzee ancestor deduced from observed gene counts in extant species.

is similarly promiscuous and has an even higher testis-weight to body-weight ratio, the rhesus MSY shows little evidence of intense sexual selection. We suggest that in the rhesus lineage, such selection was focused on spermatogenesis factors that are encoded elsewhere in the genome. This would also account for the virtual absence in rhesus of the MSY sequence amplification that is prominent in human and even more pronounced in chimpanzee (Fig. 1).

Our knowledge of all five strata of the MSY, gained through our comprehensive comparisons of ancestral gene content in the rhesus, human and chimpanzee MSYs, enabled us to reconstruct the kinetics and trajectory of human MSY evolution. For each of the five MSY strata, we estimated ancestral gene numbers at three points in the human evolutionary lineage: in the last common ancestor of human and chimpanzee (6 Myr ago), in the last common ancestor of human and rhesus macaque (25 Myr ago) and at the time of the stratum's formation, when X–Y differentiation was initiated (from ~30 to >240 Myr ago; Table 1). For each stratum, we plotted these three estimated numbers against evolutionary time, together with the observed number of ancestral genes in modern human, and fit a curve (Fig. 3 and Supplementary Fig. 18). For each of the five strata, a simple two-parameter model, using an exponential decay equation that includes a baseline constant, provides an excellent fit to our data (Fig. 3 and Supplementary Table 6). According to this reconstruction, ancestral gene decay within each stratum proceeded rapidly at first—with an ancestral gene half-life of less than 5 Myr (Supplementary Table 6)—but then decelerated markedly, as the ancestral gene count reached a stable level far below its starting point. In our reconstruction, strata 1–4 had already reached a stable level before the human lineage diverged from rhesus; after divergence from rhesus, gene loss in the human lineage was limited to stratum 5, the youngest stratum, which stabilized before the human lineage diverged from chimpanzee.

Our empirical reconstruction of MSY evolution is at odds with a linear model^{7,9,10} and with a simple random decay (exponential) model²⁵, both of which project a continual decline of MSY gene content and cannot account for the stability of gene content that we observe over the past 25 Myr (Fig. 3). Our data are better explained by more complex models for MSY gene loss that incorporate a combination of evolutionary forces²⁶. Sequencing additional Y chromosomes from animals that represent more divergent mammalian lineages will enable refinement of our reconstruction of MSY gene kinetics in the human lineage.

METHODS SUMMARY

BAC selection and sequencing. The SHIMS (single-haplotype iterative mapping and sequencing) strategy¹¹ was used to assemble a path of sequenced clones selected from the CHORI-250 BAC library (<http://bacpac.chori.org>) and a custom BAC library (RMAEX) constructed by Amplicon Express (<http://www.genomex.com>).

Fluorescence *in situ* hybridization analysis. Assays were performed on rhesus fibroblast cell line PR00112 from Coriell Institute for Medical Research (<http://ccr.coriell.org>). Extended metaphase fluorescence *in situ* hybridization (FISH) and interphase FISH were performed as previously described²⁷.

Radiation hybrid mapping. Nine sequence-tagged site (STS) markers (Supplementary Table 7) were tested on a 10,000-rad panel consisting of 185 hybrid clones²⁸. A genetic map was constructed and analysed statistically using RHMAPPER 1.22 (ref. 29).

Generation of complementary DNA for polymerase chain reaction with reverse transcription (RT–PCR) and 454 sequencing. cDNA was generated from total RNA that was isolated from male rhesus tissues using the RNeasy kit (Qiagen). For 454 sequencing, cDNA was normalized using the Trimmer kit (Evrogen).

Alignments and dot plots. Rhesus and human Y sequences were aligned using Stretcher (<http://bioweb2.pasteur.fr/docs/EMBOSS/stretcher.html>) with a gap open penalty of 20 and a gap extend penalty of 1. Dot-plot analyses were performed using custom Perl codes (http://jura.wi.mit.edu/page/papers/Hughes_et_al_2005/tables/dot_plot.pl).

Calculation of α_m . The male-to-female mutation rate ratio was calculated from the human–rhesus Y divergence rate and the human–rhesus autosomal divergence rate using a previously described method^{15,30}.

Modelling ancestral MSY gene loss. We fit a one-phase exponential decay model with a baseline constant (shown below) to our data (gene numbers shown in

Table 1) using nonlinear regression analysis in GraphPad Prism 5.0. Parameters for each stratum are given in Supplementary Table 6.

One-phase exponential decay model:

$$N(t) = (N_0 - b)e^{-Kt} + b$$

Where $N(t)$ is the number of genes at time t , N_0 is the number of genes within given stratum in ancestral autosomal or pseudoautosomal portion of genome, K is the decay constant and b is the baseline (approximated by the number of active ancestral genes within that stratum on human Y chromosome).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 October; accepted 10 January 2012.

Published online 22 February 2012.

- Charlesworth, B. The evolution of sex chromosomes. *Science* **251**, 1030–1033 (1991).
- Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
- Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
- Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
- Bellott, D. W. *et al.* Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612–616 (2010).
- Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
- Aitken, R. J. & Marshall Graves, J. A. The future of sex. *Nature* **415**, 963 (2002).
- Sykes, B. *Adam's Curse* (Norton, W. W. & Company, 2004).
- Graves, J. A., Koina, E. & Sankovic, N. How the gene content of human sex chromosomes evolved. *Curr. Opin. Genet. Dev.* **16**, 219–224 (2006).
- Marshall Graves, J. A. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu. Rev. Genet.* **42**, 565–586 (2008).
- Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29**, 279–286 (2001).
- Hughes, J. F. *et al.* Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**, 100–103 (2005).
- Hughes, J. F. *et al.* Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
- Gibbs, R. A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
- Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 863–867 (1987).
- Bohossian, H. B., Skaletsky, H. & Page, D. C. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature* **406**, 622–625 (2000).
- Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
- Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
- Lahn, B. T. & Page, D. A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. *Hum. Mol. Genet.* **9**, 311–319 (2000).
- Rozen, S., Marszałek, J. D., Alagappan, R. K., Skaletsky, H. & Page, D. C. Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying effective purifying selection. *Am. J. Hum. Genet.* **85**, 923–928 (2009).
- Dixon, A. F. *Primate Sexuality: Comparative Studies of the Prosimians, Monkeys, Apes and Human Beings* (Univ. Chicago Press, 1998).
- Rice, W. R. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167 (1987).
- Schultz, A. H. The relative weight of the testes in primates. *Anat. Rec.* **72**, 387–394 (1938).
- Harcourt, A. H., Harvey, P. H., Larson, S. G. & Short, R. V. Testis weight, body weight and breeding system in primates. *Nature* **293**, 55–57 (1981).
- Graves, J. A. Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901–914 (2006).
- Bachtrog, D. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* **179**, 1513–1525 (2008).
- Saxena, R. *et al.* The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nature Genet.* **14**, 292–299 (1996).
- Karere, G. M., Froenicke, L., Millon, L., Womack, J. E. & Lyons, L. A high-resolution radiation hybrid map of rhesus macaque chromosome 5 identifies rearrangements in the genome assembly. *Genomics* **92**, 210–218 (2008).
- Slonim, D., Kruglyak, L., Stein, L. & Lander, E. Building human genome maps with radiation hybrids. *J. Comput. Biol.* **4**, 487–504 (1997).
- Shimmin, L. C., Chang, B. H. & Li, W. H. Male-driven evolution of DNA sequences. *Nature* **362**, 745–747 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Frazzoni, G. Rogers and S. Zaghlul for technical assistance; L. Lyons, W. Murphy and J. Womack for radiation hybrid panels; W. Johnson, E. Curran, S. O'Neil and A. Kaur for tissue samples; R. Harris for analyses of rhesus–human genome alignments; T. DiCesare for assistance with figures; and D. Bellott, M. Carmell, R. Desgraz, Y. Hu, B. Lesch, J. Mueller, K. Romer and S. Soh for comments on the manuscript. This work was supported by the National Institutes of Health, the Howard Hughes Medical Institute and the Charles A. King Trust.

Author Contributions J.F.H., H.Sk., W.C.W., S.R., R.A.G., R.K.W. and D.C.P. planned the project. J.F.H., H.Sk., L.G.B., T.J.C. and N.K. performed BAC mapping, radiation hybrid mapping and real-time polymerase chain reaction analyses. T.G., R.S.F., S.D., Y.D., C.J.B.,

C.K., Q.W., H.Sh., M.H., D.V., L.V.N., A.C., L.C., J.V., H.K. and D.M.M. were responsible for BAC sequencing. J.F.H. and H.Sk. performed comparative sequence analyses. T.P. performed FISH analyses. J.F.H. and D.C.P. wrote the paper.

Author Information cDNA sequences of rhesus Y genes have been deposited in GenBank (<http://www.ncbi.nlm.nih.gov>) under accession numbers FJ527009–FJ527028 and FJ648737–FJ648739. 454 testis cDNA sequences have been deposited in GenBank under accession number SRA039857. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.F.H. (jhughes@wi.mit.edu).

METHODS

BAC selection and sequencing. The SHIMS (single-haplotype iterative mapping and sequencing) strategy¹¹ was used to assemble a path of sequenced clones selected from the CHORI-250 BAC library (<http://bacpac.chori.org>) and a custom BAC library (RMAEX) constructed by Amplicon Express (<http://www.genomex.com>). The rate of error in the finished sequence was estimated by counting mismatches in overlapping clones.

FISH analysis. All assays were performed on rhesus fibroblast cell line PR00112 obtained from the Coriell Institute for Medical Research (<http://ccr.coriell.org>). Extended metaphase FISH and interphase FISH were performed as previously described²⁷.

Radiation hybrid mapping. Nine STS markers (Supplementary Table 7) were tested on a 10,000-rad, male whole-genome panel consisting of 185 hybrid clones²⁸. The average retention frequency of the markers tested was 16%, ranging from 10–27%. A genetic map was constructed and analysed statistically using RHMAPPER 1.22 (ref. 29).

RT-PCR. Total RNA was isolated from male rhesus tissues (brain, prostate, liver, lung and spleen testis; Alpha Genesis) using the RNeasy kit (Qiagen) and cDNA was generated. RT-PCR primer sequences and product sizes are listed in Supplementary Table 8.

454 sequencing of testis cDNA. Rhesus testis cDNA was generated from total RNA isolated using the RNeasy kit (Qiagen). The cDNA was normalized using the Trimmer kit (Evrogen) and sequenced on a 454 FLX Titanium machine.

Alignments and dot plots. Rhesus and human Y sequences were aligned using Stretcher (<http://bioweb2.pasteur.fr/docs/EMBOSS/stretcher.html>) with a gap open

penalty of 20 and a gap extend penalty of 1. Dot plot analyses were performed using custom Perl codes (http://jura.wi.mit.edu/page/papers/Hughes_et_al_2005/tables/dot_plot.pl).

Calculation of α_m . The male-to-female mutation rate ratio was calculated using the human–rhesus Y divergence rate (9.40%, 312,840 substitutions per 3,330,847 sites examined) and the human–rhesus autosomal divergence rate (1.385×10^8 substitutions per 2.248×10^{10} sites examined; hg18-rheMac2 alignments downloaded from <http://www.genome.ucsc.edu>). Miyata's formula was then used to calculate α_m (refs 15, 30): $Y/A = 2\alpha_m/(1 + \alpha_m)$. Confidence intervals for ratios of divergence rates were calculated as previously described³⁰.

Modelling ancestral MSY gene loss. We modelled the numbers of ancestral genes within individual MSY strata as a function of time in millions of years before the present by fitting a one-phase exponential decay model with a baseline constant (below) to our data (gene numbers shown in Table 1) using nonlinear regression analysis in GraphPad Prism 5.0. Parameters for each stratum are given in Supplementary Table 6.

One-phase exponential decay model:

$$N(t) = (N_0 - b)e^{-Kt} + b$$

Where $N(t)$ is the number of genes at time t , N_0 is the number of genes within a given stratum in the ancestral autosomal or pseudoautosomal portion of genome, K is the decay constant and b is the baseline (approximated by the number of active ancestral genes within that stratum on human Y chromosome).

Repetitive motor learning induces coordinated formation of clustered dendritic spines *in vivo*

Min Fu¹, Xinzhu Yu¹, Ju Lu² & Yi Zuo¹

Many lines of evidence suggest that memory in the mammalian brain is stored with distinct spatiotemporal patterns^{1,2}. Despite recent progresses in identifying neuronal populations involved in memory coding^{3–5}, the synapse-level mechanism is still poorly understood. Computational models and electrophysiological data have shown that functional clustering of synapses along dendritic branches leads to nonlinear summation of synaptic inputs and greatly expands the computing power of a neural network^{6–10}. However, whether neighbouring synapses are involved in encoding similar memory and how task-specific cortical networks develop during learning remain elusive. Using transcranial two-photon microscopy¹¹, we followed apical dendrites of layer 5 pyramidal neurons in the motor cortex while mice practised novel forelimb skills. Here we show that a third of new dendritic spines (post-synaptic structures of most excitatory synapses) formed during the acquisition phase of learning emerge in clusters, and that most such clusters are neighbouring spine pairs. These clustered new spines are more likely to persist throughout prolonged learning sessions, and even long after training stops, than non-clustered counterparts. Moreover, formation of new spine clusters requires repetition of the same motor task, and the emergence of succedent new spine(s) accompanies the strengthening of the first new spine in the cluster. We also show that under control conditions new spines appear to avoid existing stable spines, rather than being uniformly added along dendrites. However, succedent new spines in clusters overcome such a spatial constraint and form in close vicinity to neighbouring stable spines. Our findings suggest that clustering of new synapses along dendrites is induced by repetitive activation of the cortical circuitry during learning, providing a structural basis for spatial coding of motor memory in the mammalian brain.

Spines are dendritic protrusions that contain all the essential components for postsynaptic signalling and are thus a good indicator of synaptic connectivity^{12,13}. The clustered plasticity model suggests that neighbouring spines tend to transmit similar information to the postsynaptic neuron^{6,7}. To investigate the formation and functional significance of spine clusters during learning, we trained *thy1-YFP-H* mice¹⁴ with a seed-reaching task¹⁵ and followed the dynamics of spines on apical dendrites of layer 5 (L5) pyramidal neurons in the motor cortex contralateral to the trained limb during different learning phases. We found that $32.5 \pm 2.2\%$ of new spines that formed during the acquisition phase of learning (early training, days 1–4) emerged in clusters; that is, two or more neighbouring new spines without interspersed existing spine(s) (Fig. 1a, b). Most such clusters (61 cases) comprised two contiguous new spines, and the other two clusters comprised three. In contrast, fewer new spine clusters emerged in untrained control mice over the same period ($6.8 \pm 4.6\%$, $P < 0.01$) or in trained mice during the consolidation phase of learning (late training, days 13–16; $7.4 \pm 4.3\%$, $P < 0.01$; Fig. 1b). In addition to clustering of contiguous new spines, we observed a few cases in which two or more new spines formed in close vicinity to each other, but with

up to three existing spines interspersed among them, as well as cases in which new filopodia clustered with new spines (Supplementary Fig. 1). We incorporated these cases in another set of analyses, in which a cluster was defined as a set of new spines/filopodia formed within $5 \mu\text{m}$ of each other, regardless of the presence or absence of existing spine(s) between them (Supplementary Information). These analyses again revealed that a significantly higher percentage of new spines clustered during early training, compared with that in controls or during late training ($P < 0.01$ for both cases; Supplementary Fig. 2). They also showed that filopodia only made a minor contribution to new protrusion clusters (Supplementary Fig. 3). More interestingly, among the new spines observed at the end of the acquisition phase (day 4), clustered new spines had a significantly higher survival rate than non-clustered ones (that is, individual new spines flanked by two existing spines) by training day 16 ($P < 0.01$), as well as 4 months after training stopped ($P < 0.05$; Fig. 1c). Together, our results reveal that motor learning induces coordinated formation of clustered synapses, which presumably belong to the same neuronal circuit and persist over time to encode motor information.

Perfection of a motor skill requires repeated practice, usually through multiple training sessions. We therefore sought to find out whether clustered new spines observed on training day 4 were formed within the same training session or across different sessions. We imaged the mice three times (on the day before training, and after 1 and 4 days of training), and found that among new spine clusters observed on training day 4, only 2.4% were composed of spines that formed together between training days 0 and 1. On the other hand, 43.9% of clusters were composed of spines formed between days 1 and 4, and the remaining 53.7% of clusters consisted of one spine formed between days 0 and 1 (the first new spine) and another spine formed between days 1 and 4 (the second new spine). Thus most new spine clusters emerged through recurrent training sessions. To determine how the formation of the second new spine in a cluster correlates with functional changes of the first new spine, we categorized first new spines into three groups based on their survival and neighbouring spine addition: transient new spines (formed on training day 1 but lost by day 4); persistent clustered new spines (formed on training day 1, survived until day 4, with an adjacent new spine formed between days 1 and 4); and persistent non-clustered new spines (formed on training day 1, survived until day 4, with no adjacent new spine formation) (Fig. 1d). As spine head size closely correlates with synaptic strength, we followed head sizes of first new spines over time. On training day 1, we found that head sizes of both transient and persistent new spines were significantly smaller than those of existing stable spines along the same dendrite ($P < 0.001$ for both cases, Supplementary Fig. 4). By training day 4, head sizes of persistent clustered new spines increased significantly ($P < 0.01$; Fig. 1e and Supplementary Fig. 5a), whereas head sizes of persistent non-clustered new spines remained comparable to day 1 ($P > 0.2$; Fig. 1f and Supplementary Fig. 5b). Because spine head size is a good proxy for synaptic strength, these data suggest that formation of the second new spines

¹Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, California 95064, USA. ²Department of Biological Sciences and James H. Clark Center, Stanford University, Stanford, California 94305, USA.

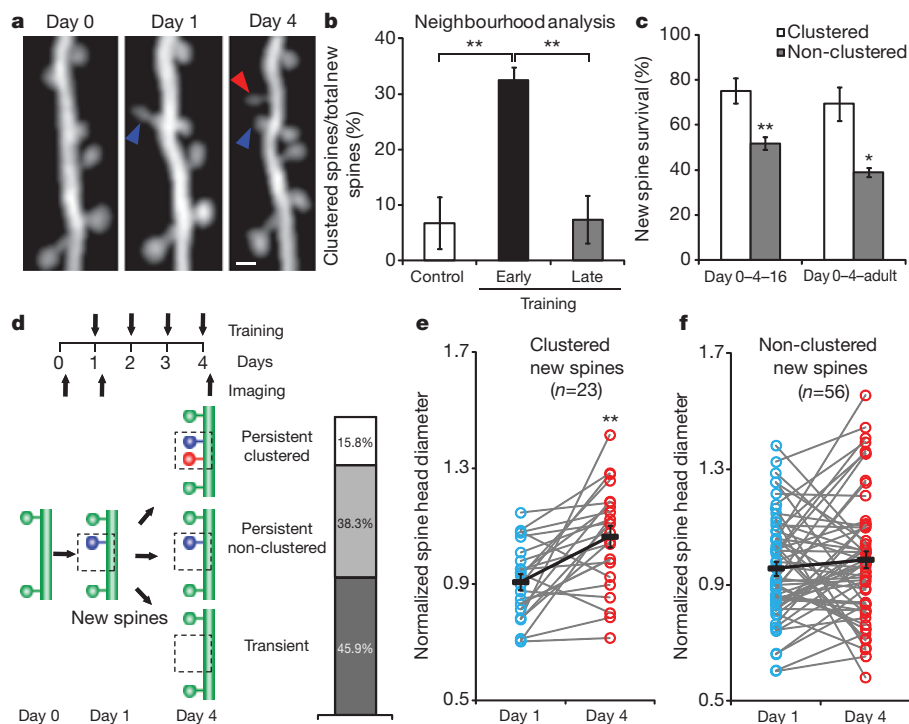


Figure 1 | Acquisition of a novel motor skill induces formation of spine clusters. **a**, Repeated imaging of the same dendritic branch during motor learning reveals that a second new spine that formed between days 1 and 4 (red arrowhead) is located next to a stabilized new spine that had formed on day 1 (blue arrowhead). Scale bar, 1 μ m. **b**, A higher percentage of new spines formed in clusters over 4 days during early training ($n = 18$ mice), compared with control ($n = 7$) and late training ($n = 4$). **c**, Clustered new spines observed on training day 4 have a higher survival rate than non-clustered counterparts by the end of the 16-day training ($n = 6$), as well as 4 months after training stops ($n = 4$). **d**, New spines formed on training day 1 are classified according to their fate and neighbouring spine formation. **e**, Spine head sizes of persistent clustered new spines increase between training days 1 and 4. **f**, Spine head sizes of persistent non-clustered new spines show no change between training days 1 and 4. Spine head size is quantified by the normalized spine head diameter, defined as the ratio of the spine head diameter to the adjacent dendritic shaft diameter. * $P < 0.05$, ** $P < 0.01$. Error bars, s.e.m.

accompanies synaptic potentiation at first new spines during motor learning. It is consistent with previous reports that long-term potentiation at a single spine can facilitate formation and potentiation of neighbouring spines^{16,17}.

Different sets of synapses have been shown to be involved in different motor tasks¹⁵. We therefore trained the same mice sequentially with two motor skills (cross-training) to determine if spines induced by different motor tasks cluster. Cross-training started with the reaching task on day 1 and then switched to the capellini-handling task, which also requires forelimb coordination, on days 2–4 (Fig. 2a and Supplementary Table 1). We found that $12.3 \pm 0.4\%$ new spines formed during the capellini-handling task between days 1 and 4, among which $28.4 \pm 2.8\%$ occurred in clusters (Fig. 2b–d). Both the spine formation rate and the percentage of clustered new spines were comparable to those in mice continuously trained with the reaching task (reach-only) ($P > 0.5$ in both cases), and were significantly higher than those in control mice over the same period of time ($P < 0.01$ in both cases, Fig. 2c, d). Thus, the capellini-handling task itself can induce clustered spine formation. However, only $3.3 \pm 2.1\%$ of capellini-handling-induced new spines clustered with reaching-induced new spines in cross-training. This contrasts with the outcome of reach-only training ($13.8 \pm 1.0\%$, $P < 0.01$, Fig. 2e), suggesting that new spines induced by different tasks have a low incidence of clustering with each other. To characterize further the task-specific nature of clustered spine formation, we housed animals in a motor enriched environment with daily change of motor challenges (Fig. 2a; see Methods). Motor enrichment also robustly enhanced spinogenesis: $13.7 \pm 0.8\%$ new spines formed between days 1 and 4, comparable to the percentages under reach-only and cross-training conditions ($P > 0.1$ for both cases). However, only $12.6 \pm 1.1\%$ of these new spines appeared in clusters, a percentage comparable to controls ($P > 0.2$, Fig. 2d) but significantly lower than that under reach-only or cross-training conditions ($P < 0.01$ for both cases). Together these data indicate that, whereas novelty in learning stimulates spinogenesis, repetitive activation of the same cortical circuit is crucial in clustered spine formation (Supplementary Fig. 6).

The phenomenon of learning-induced, coordinated spinogenesis led us to investigate further the spatial distribution of new spines. We first examined the distance between a new spine (n) and its nearest existing

spine (s) (D_{n-s} , Fig. 3a) in control mice. We then simulated D_{n-s} distribution under the null hypothesis that new spines form uniformly and independently along the dendrite (see Methods). Compared with simulation results, the median of measured values of D_{n-s} was significantly larger (Fig. 3b), and the cumulative probability distribution of measured values of D_{n-s} was shifted towards longer distances (Fig. 3c). These results suggest that new spines are not randomly dispersed along dendritic segments, and their apparent avoidance of existing stable spines under control conditions is consistent with the idea that neighbouring spines share and compete for local resources^{18–21}.

To determine if motor learning alters the spatial distribution of new spines, we examined values of D_{n-s} in mice trained with the reaching task. We found that the distance between a new spine formed on training day 1 (n_1) and the nearest existing spine (D_{n_1-s}) was comparable for trained and control mice ($P > 0.7$). We classified new spines formed between training days 1 and 4 (n_2) into two categories: clustered n_2 (that is, those that formed next to a stabilized first new spine; Fig. 3d) and non-clustered n_2 (those that did not form next to a stabilized first new spine). We found that clustered n_2 were significantly closer to their nearest stable spines (n_1 or a stable spine existing since day 0; Fig. 3d) ($D_{n_2-s, \text{clustered}}$) than were n_1 (D_{n_1-s}) ($P < 0.05$). In contrast, the distance between a non-clustered n_2 and its nearest stable spine ($D_{n_2-s, \text{non-clustered}}$) was comparable to D_{n_1-s} ($P > 0.9$, Fig. 3e). In addition, when an n_1 formed between two adjacent stable spines, the distance between the two stable spines (D_{s-n_1-s}) was comparable for control and trained mice ($P > 0.7$; Supplementary Fig. 7). However, the distance between a stabilized n_1 and the adjacent stable spine, between which a clustered n_2 formed ($D_{n_1-n_2-s}$), was significantly smaller than the distance between two adjacent stable spines, between which a non-clustered n_2 formed (D_{s-n_2-s} ; $P < 0.01$; Supplementary Fig. 7). These results suggest that learning-induced clustered new spines can overcome the spatial constraint of existing spines and be packed into tighter dendritic space.

Recent studies have shown that dendritic spines are dynamic in the living brain, and that rearrangement of cortical connections through *de novo* growth and loss of spines provides a structural substrate for experience-dependent plasticity^{22–25}. Built upon these works, our study reveals a novel spatial rule of spinogenesis during motor learning. We found that learning-induced new spines tend to form in small clusters (mostly pairs). The correlation between the emergence of the second

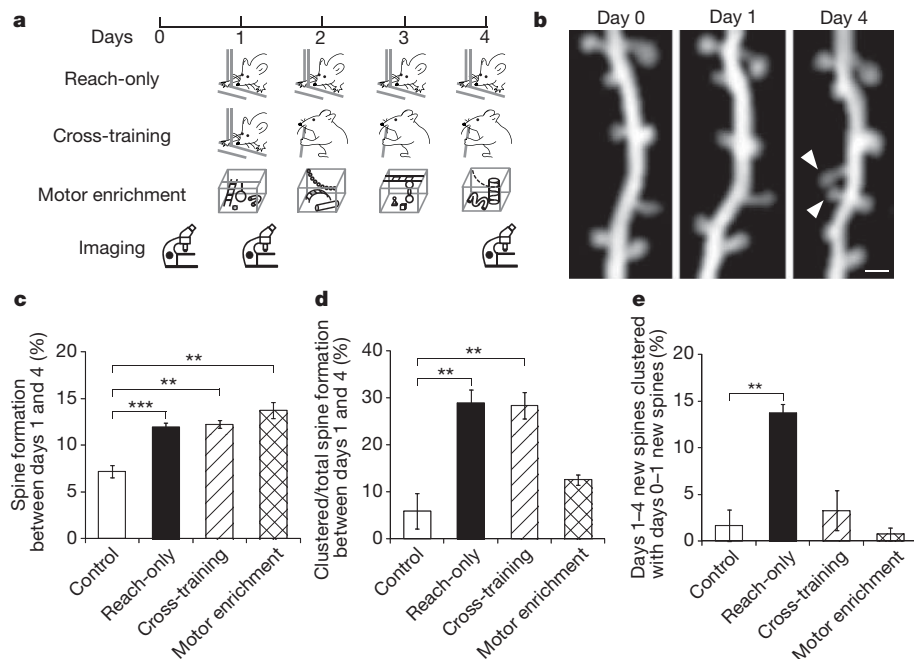


Figure 2 | Clustered new spines form over multiple training sessions of the same, but not different, motor tasks. **a**, Timelines of reach-only, cross-training and motor enrichment experiments. **b**, Repeated imaging of the same dendritic branch revealed that two neighbouring new spines (arrowheads) formed between days 1 and 4 during cross-training. Scale bar, 1 μm . **c**, Higher percentages of new spines formed between days 1 and 4 in reach-only, cross-training and motor enrichment, compared with controls. **d**, Higher percentages

of new spines formed in clusters between days 1 and 4 in reach-only and cross-training, compared with controls. **e**, A higher percentage of new spines that formed between days 1 and 4 clustered with new spines that had formed between days 0 and 1 in the reach-only condition, compared with controls. Number of mice examined: six control, nine reach-only, five cross-training and six motor enrichment. $**P < 0.01$, $***P < 0.001$. Error bars, s.e.m.

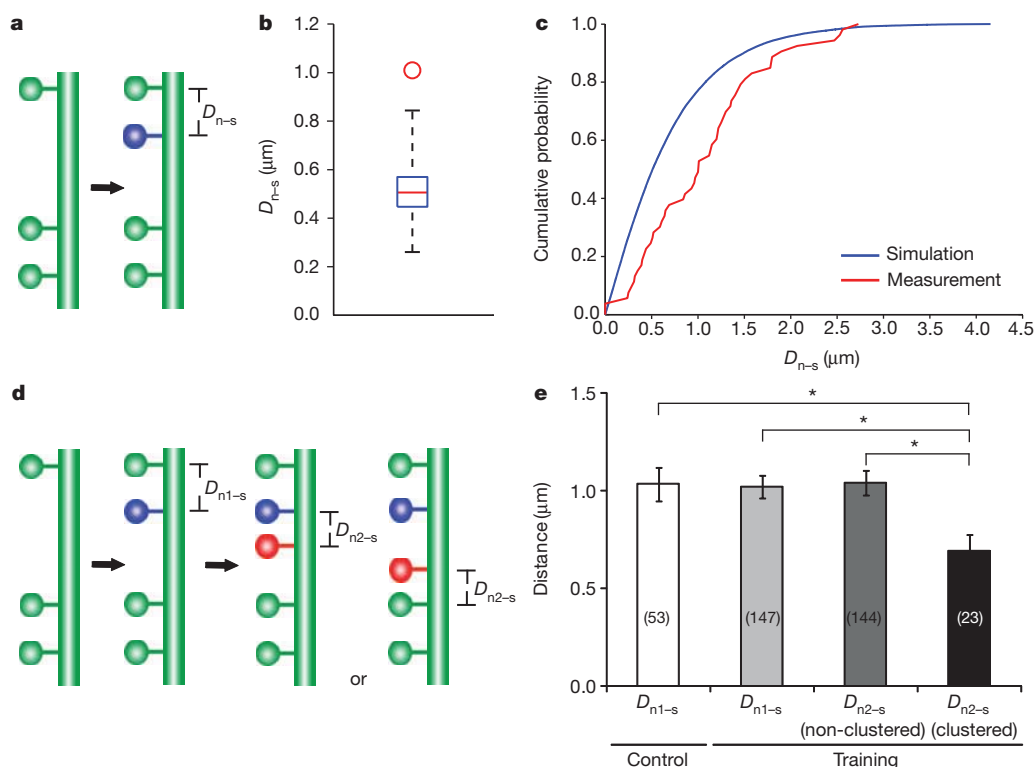


Figure 3 | The spatial distribution of new spines along dendrites.

a, Schematic illustrating the measurement of D_{n-s} . **b**, The median of measured values of D_{n-s} (red circle) is significantly larger than that of simulated values of D_{n-s} (box plot of results from 1,000 simulations, with whiskers representing the minimum and the maximum) in control mice. The simulation is based on the null hypothesis that new spines are added independently and uniformly along a linear dendrite. **c**, Cumulative probability distribution of measured D_{n-s} is shifted towards longer distances than the simulated D_{n-s} in control mice.

d, Schematic illustrating the measurement of $D_{n2-s, \text{clustered}}$. The nearest spine to a clustered n_2 could be either a persistent first new spine (n_1) or a stable spine existing since day 0, depending on relative n_2 location. **e**, D_{n1-s} in control mice is comparable to that of trained mice. In trained mice, $D_{n2-s, \text{clustered}}$ is significantly smaller than D_{n1-s} , whereas $D_{n2-s, \text{non-clustered}}$ is comparable to D_{n1-s} . The number of spines analysed in each condition is indicated on each column. $*P < 0.05$. Error bars, s.e.m.

new spine and the strengthening of the first new spine also suggests their potential participation in the same neuronal circuit. These findings support the clustered plasticity model, which postulates that synapses located close together along the same dendritic branch are more likely to be allocated for the same information than synapses dispersed throughout the dendritic arbor⁷. Indeed, in the mouse auditory cortex, although spines tuned for different frequencies are highly interspersed, 26% of neighbouring spines exhibit similar effective frequencies, much more frequently than anticipated from random distribution (10%)²⁶. Therefore, although neurons tend to maximize their overall connections²⁷, clustered plasticity ensures strengthening of circuit-specific connections and enables spatial coding for task-related information.

Previous electron microscopy studies have revealed that neighbouring spines can form synapses with the same axon^{28–30} (see Supplementary Fig. 8a, c, e). Positioning multiple synapses between a pair of neurons in close proximity allows nonlinear summation of synaptic strength, and potentially increases the dynamic range of synaptic transmission well beyond what can be achieved by random positioning of the same number of synapses. Alternatively, clustered new spines may synapse with distinct (but presumably functionally related) presynaptic partners (Supplementary Fig. 8b, d). In this case, they could potentially integrate inputs from different neurons nonlinearly and increase the circuit's computational power. Distinguishing between these two possibilities would probably require circuit reconstruction by electron microscopy following *in vivo* imaging to reveal the identities of presynaptic partners of newly formed spines.

Profiling spine formation during novel experiences, our data revealed a critical role of repetitive activation of the same neuronal circuit. The fact that the second new spine in a cluster can overcome the spatial constraint imposed by existing spines suggests that repetitive activation of a neuronal circuit may modify or reallocate local 'resources' for spinogenesis. Such resources may be permissive or instructive molecular cues at the pre- or postsynaptic site, or the availability of suitable partners (for example, axonal boutons). Understanding the nature and regulation of such resources may hold the key to elucidating the cellular mechanisms of clustered spine formation. It will be conducive to the development of tools to label and manipulate specific synaptic populations, and ultimately to the dissection of the causal relationship between synaptic dynamics and learning.

METHODS SUMMARY

YFP-H line mice¹⁴ expressing yellow fluorescent protein (YFP) in a small subset of cortical neurons were used in all the experiments. Mice of both sexes were trained with different motor-skill tasks or housed in a motor-enriched environment, starting at 1 month of age (see Methods). The procedures for transcranial two-photon imaging and quantification of spine dynamics have been described previously^{11,15}. ImageJ was used to measure spine head size, as well as inter-spine distances. Simulation was performed with custom-written codes in Matlab (MathWorks) and statistical analyses were performed using GraphPad Prism 5 (GraphPad Software) (see Methods). All data were presented as mean \pm s.e.m. *P* values were calculated using the Mann–Whitney *U*-test for independent samples, and the Wilcoxon signed-rank test for paired samples.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 September 2011; accepted 9 January 2012.

Published online 19 February 2012.

1. Silva, A. J., Zhou, Y., Rogerson, T., Shobe, J. & Balaji, J. Molecular and cellular approaches to memory allocation in neural circuits. *Science* **326**, 391–395 (2009).
2. Aimone, J. B., Wiles, J. & Gage, F. H. Potential role for adult neurogenesis in the encoding of time in new memories. *Nature Neurosci.* **9**, 723–727 (2006).

3. Han, J. H. *et al.* Selective erasure of a fear memory. *Science* **323**, 1492–1496 (2009).
4. Komiya, T. *et al.* Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464**, 1182–1186 (2010).
5. Kee, N., Teixeira, C. M., Wang, A. H. & Frankland, P. W. Preferential incorporation of adult-generated granule cells into spatial memory networks in the dentate gyrus. *Nature Neurosci.* **10**, 355–362 (2007).
6. Larkum, M. E. & Nevian, T. Synaptic clustering by dendritic signalling mechanisms. *Curr. Opin. Neurobiol.* **18**, 321–331 (2008).
7. Govindarajan, A., Kelleher, R. J. & Tonegawa, S. A clustered plasticity model of long-term memory engrams. *Nature Rev. Neurosci.* **7**, 575–583 (2006).
8. Poirazi, P. & Mel, B. W. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* **29**, 779–796 (2001).
9. Sjostrom, P. J. & Hausser, M. A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron* **51**, 227–238 (2006).
10. Losonczy, A. & Magee, J. C. Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron* **50**, 291–307 (2006).
11. Zuo, Y., Lin, A., Chang, P. & Gan, W. B. Development of long-term dendritic spine stability in diverse regions of cerebral cortex. *Neuron* **46**, 181–189 (2005).
12. Segal, M. Dendritic spines and long-term plasticity. *Nature Rev. Neurosci.* **6**, 277–284 (2005).
13. Harms, K. J. & Dunaevsky, A. Dendritic spine plasticity: looking beyond development. *Brain Res.* **1184**, 65–71 (2007).
14. Feng, G. *et al.* Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
15. Xu, T. *et al.* Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* **462**, 915–919 (2009).
16. Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
17. De Roo, M., Klausner, P. & Muller, D. LTP promotes a selective long-term stabilization and clustering of dendritic spines. *PLoS Biol.* **6**, e219 (2008).
18. Gray, N. W., Weimer, R. M., Bureau, I. & Svoboda, K. Rapid redistribution of synaptic PSD-95 in the neocortex *in vivo*. *PLoS Biol.* **4**, e370 (2006).
19. Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321**, 136–140 (2008).
20. Rose, J., Jin, S. X. & Craig, A. M. Heterosynaptic molecular dynamics: locally induced propagating synaptic accumulation of CaM kinase II. *Neuron* **61**, 351–358 (2009).
21. Tsurui, S. *et al.* Local sharing as a predominant determinant of synaptic matrix molecular dynamics. *PLoS Biol.* **4**, e271 (2006).
22. Yu, X. & Zuo, Y. Spine plasticity in the motor cortex. *Curr. Opin. Neurobiol.* **21**, 169–174 (2011).
23. Chen, J. L. & Nedivi, E. Neuronal structural remodeling: is it all about access? *Curr. Opin. Neurobiol.* **20**, 557–562 (2010).
24. Fu, M. & Zuo, Y. Experience-dependent structural plasticity in the cortex. *Trends Neurosci.* **34**, 177–187 (2011).
25. Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Rev. Neurosci.* **10**, 647–658 (2009).
26. Chen, X., Leischner, U., Rochefort, N. L., Nelken, I. & Konnerth, A. Functional mapping of single spines in cortical neurons *in vivo*. *Nature* **475**, 501–505 (2011).
27. Yuste, R. Dendritic spines and distributed circuits. *Neuron* **71**, 772–781 (2011).
28. Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E. & Svoboda, K. Spine growth precedes synapse formation in the adult neocortex *in vivo*. *Nature Neurosci.* **9**, 1117–1124 (2006).
29. Fiala, J. C., Allwardt, B. & Harris, K. M. Dendritic spines do not split during hippocampal LTP or maturation. *Nature Neurosci.* **5**, 297–298 (2002).
30. Toni, N., Buchs, P. A., Nikonenko, I., Bron, C. R. & Muller, D. LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature* **402**, 421–425 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. States, D. Garcia, L. Hinck, T. Jones, S. Song, W. Thompson and G. Wang for comments on this manuscript. We thank A. Perlik and T. Xu for technical support. This work was supported by grants from the DANA Foundation and the National Institutes of Mental Health to Y.Z.

Author Contributions M.F. and X.Y. did the *in vivo* imaging and made the figures. M.F. performed behavioural training and all spine analyses, and made figures for repetitive imaging. J.L. and M.F. performed Matlab simulation and statistical analyses. J.L., M.F. and X.Y. participated in discussion about the paper. Y.Z. initiated and designed the project, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.Z. (yizuo@ucsc.edu).

METHODS

Experimental animals. *Thy1-YFP-H* line mice¹⁴ were obtained from the Jackson Laboratory. Mice were group-housed and bred in the University of California, Santa Cruz, animal facility, with all experiments performed in accordance with approved animal protocols.

Motor skill training and motor enrichment. Both the mouse single-seed reaching task and capellini-handling task protocols have been previously described¹⁵. 'Motor-enriched' mice were reared in groups of 8–12 in large cages (90 cm × 25 cm × 15 cm) containing various toys, such as ropes, ladders, chains, hanging mesh/bars etc., all of which required substantial motor coordination. The nature of toys was changed on a daily basis. Control mice were housed in standard mouse cages, with up to five mice per cage.

Surgical procedure for *in vivo* transcranial imaging. The procedure for transcranial two-photon imaging has been described previously^{11,31}. Trained mice were imaged immediately after each training session.

Data quantification. All analyses of spine dynamics were done using ImageJ software, blinded for experimental conditions. Quantification criteria of dendritic spines have been described previously¹⁵. All dendritic protrusions were tracked manually in three-dimensional stacks to ensure the consistency of protrusion identification across imaging sessions, despite possible tissue movement or rotation. The number and location of dendritic protrusions (defined as protrusion length larger than one-third of dendritic shaft diameter) were identified in each view. Filopodia were identified as long, thin structures with the ratio of head diameter to neck diameter being less than 1.2 and the ratio of length to neck diameter being greater than 3. The remaining protrusions were classified as spines. Formation and elimination of spines and filopodia were determined by comparing images collected at two different time points. Spines or filopodia were considered identical between the two images if they were within 0.7 μm of their expected positions, based on their spatial relationship to adjacent landmarks and/or their positions relative to immediately adjacent spines. A stable spine was defined as a spine that was present in both images. A new spine was a spine that appeared in a subsequent image but was absent from the initial image. Percentages of formed and eliminated spines (or dendritic protrusions) were normalized to the number of spines (or dendritic protrusions) in the initial image. Spine diameter analyses have been previously

described¹¹. Because imaging and animal conditions varied over time, the ratio of the spine head diameter to the adjacent dendritic shaft diameter was used as the normalized spine head diameter. Measurement of spine head intensity, as described previously³², was also performed to confirm these spine size results. Briefly, we determined the signal intensity (defined as the sum intensity of all pixels composing the spine in the best focal plane) and subtracted the background intensity (defined as the sum intensity of a region composed of the same number of pixels as the spine but with no YFP-labelled structure). The difference was then divided by the mean intensity of the adjacent dendritic shaft (defined similarly as the difference between the mean signal intensity of the shaft and the mean background intensity) to correct for varying imaging conditions. The final value was termed 'integrated spine brightness.' All distance measurements were done in ImageJ. To simulate spine formation, we first obtained the relative location of stable spines by measuring inter-spine distances along traced dendrites in seven control animals, and concatenated dendritic segments from each animal into a single 'synthetic dendrite.' We then used custom-written Matlab codes to simulate the addition of new spines. As we observed, two spines can extend from the same linear location along the dendritic segment and point towards different directions, given the cylindrical shape of dendrites. In our analysis and simulation, we made the simplifying approximation that the dendritic segment is one-dimensional rather than a tube. Therefore, zero inter-spine distance in our analysis represents two spines overlapping in linear position but actually located at different sites around the circumference of the dendritic segment. In each round of simulation, the same numbers of new spines as observed in experiments were generated independently and uniformly along synthetic dendrites. The distance between each new spine and its nearest stable spine (D_{n-s}) was calculated. The simulation was repeated 1,000 times and the resultant data were pooled to compute the simulated sample median and the cumulative probability curve. All data were presented as mean \pm standard error of mean (s.e.m.). *P* values were calculated using the Mann–Whitney *U*-test for independent samples, and the Wilcoxon signed-rank test for paired samples.

31. Grutzendler, J., Kasthuri, N. & Gan, W. B. Long-term dendritic spine stability in the adult cortex. *Nature* **420**, 812–816 (2002).
32. Hofer, S. B., Mrsic-Flogel, T. D., Bonhoeffer, T. & Hubener, M. Experience leaves a lasting structural trace in cortical circuits. *Nature* **457**, 313–317 (2009).

Continental collision slowing due to viscous mantle lithosphere rather than topography

Marin Kristen Clark¹

Because the inertia of tectonic plates is negligible, plate velocities result from the balance of forces acting at plate margins and along their base¹. Observations of past plate motion derived from marine magnetic anomalies provide evidence of how continental deformation may contribute to plate driving forces^{2–8}. A decrease in convergence rate at the inception of continental collision is expected because of the greater buoyancy of continental than oceanic lithosphere^{2,3}, but post-collisional rates are less well understood. Slowing of convergence has generally been attributed to the development of high topography that further resists convergent motion^{7–10}; however, the role of deforming continental mantle lithosphere on plate motions has not previously been considered. Here I show that the rate of India's penetration into Eurasia has decreased exponentially since their collision. The exponential decrease in convergence rate suggests that contractional strain across Tibet has been constant throughout the collision at a rate of $7.03 \times 10^{-16} \text{ s}^{-1}$, which matches the current rate. A constant bulk strain rate of the orogen suggests that convergent motion is resisted by constant average stress (constant force) applied to a relatively uniform layer or interface at depth. This finding follows new evidence that the mantle lithosphere beneath Tibet is intact¹¹, which supports the interpretation that the long-term strain history of Tibet reflects deformation of the mantle lithosphere. Under conditions of constant stress and strength, the deforming continental lithosphere creates a type of viscous resistance that affects plate motion irrespective of how topography evolved.

Marked slowing of the convergence rate of India with Eurasia determined from plate circuit motions^{2,3} agrees broadly with the estimate of continental collision age from the stratigraphic record¹², which is most simply interpreted as a response to the increased buoyancy of continental crust entering the subduction zone after collision. However, updated motions based on greater detail of spreading in the Indian Ocean¹³, motion between Nubia and North America⁴ and between North America and Eurasia¹⁴ reveal further slowing of India–Eurasia motion in late Cenozoic time as post-collisional convergence between India and Eurasia continued^{7,8}. Decreases in rate of more than 40% between 20 and 10 Myr ago occurred in the same time interval as the onset of east–west extension in the central plateau, increased faulting and erosion rates on the perimeter of the plateau, and deformation of the Indian oceanic lithosphere⁷. The correlation of a decrease in rate with tectonic activity along the margins of the plateau has been interpreted as an isostatic response to the removal of dense mantle lithosphere from beneath Tibet and an attendant increase in elevation leading to slowing of India's convergence⁷. Dynamic models constrained by modern kinematic data also conclude that gravitational potential energy (GPE) contributes stresses approximately equal in magnitude to plate boundary stresses and thus contributes significantly to the style and magnitude of deformation within Tibet¹⁵.

The potential role of topography in post-collisional slowing, however, is not confined to just the past 20 Myr. Rather, convergence rates have been exponentially decreasing since collision. Here I show how the continual slowing of India's convergence with Eurasia can be related

to the horizontal extent of deformation rather than to a period of rapid increase in elevation. On the basis of geological studies, the location of compressional faulting describes the extent over which the convergent motion is distributed. Hence, given a knowledge of speed and distance, changes in the post-collisional convergence rate can be directly interpreted as bulk continental strain rate through time and can be used to evaluate whether or not topography influences plate motion.

Before collision, the southern Eurasian margin developed as an Andean-style mountain range with abundant magmatic activity and a back-arc fold and thrust belt that significantly shortened and probably elevated the crust¹⁶. However, to the north and northeast, broad areas of non-marine deposition occurred and most researchers think that little crustal thickening occurred over Eurasia's interior¹⁷. Near the time of collision at ~45–50 Myr ago, northeast–southwest shortening reached far northward to include a much larger area of the Eurasian continent than it did previously. In fact, early faulting is now recognized across much of the modern northern boundary of the plateau (~40° N), once thought to be the most youthful part of the orogen. On the basis of new ages of surface faulting from stratigraphy and thermochronometry studies, widespread deformation at the modern northern plateau boundary occurred in Palaeocene and Eocene time^{18–23} (Fig. 1). This deformation occurred on major thrust and reverse faults that were spatially continuous for several tens to hundreds of kilometres. Faults accommodated north–south to northeast–southwest directed shortening and were large enough to develop foreland basins²³, which suggests thrust loading of sufficient magnitude to depress the Earth's surface. Collision-age deformation has occurred elsewhere in the northern interior of the plateau^{21,24,25} such that the northern plateau boundary can be considered to be a stationary feature since roughly collision time and that the complete north–south extent of the plateau probably deforms by compressional or related faulting soon after collision.

The stationary position of the northern plateau boundary relative to stable Eurasia since collision presents a unique opportunity to derive the history of orogen width on the basis of convergent motion and thus relate strain rate to changes in convergence rates through time. I define the modern orogen width (L_0) as the average extent of the plateau in the direction of plate convergence. To describe the orogen width through time $L(t)$, I use the distance between a point on the northern edge of the undeformed Indian plate (x) and the northern plateau boundary. Changes in orogen width through time can be expressed by

$$L(t) = x(t) + L_0 \quad (1)$$

where $x(t)$ represents the position of India at some time relative to its modern position and $x(t) = 0$ at $t = 0$, and positions in the past are considered to be negative time (Fig. 1). $L(t)$ decreases towards the present as India moves northward towards a stationary boundary. This description of orogen width approximates the one-dimensional length scale of continental deformation in the direction of plate convergence and averages along strike differences in the distance away from the plate boundary that deformation occurs. Changes in orogen width (L) could include processes such as the partial subduction or

¹Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA.

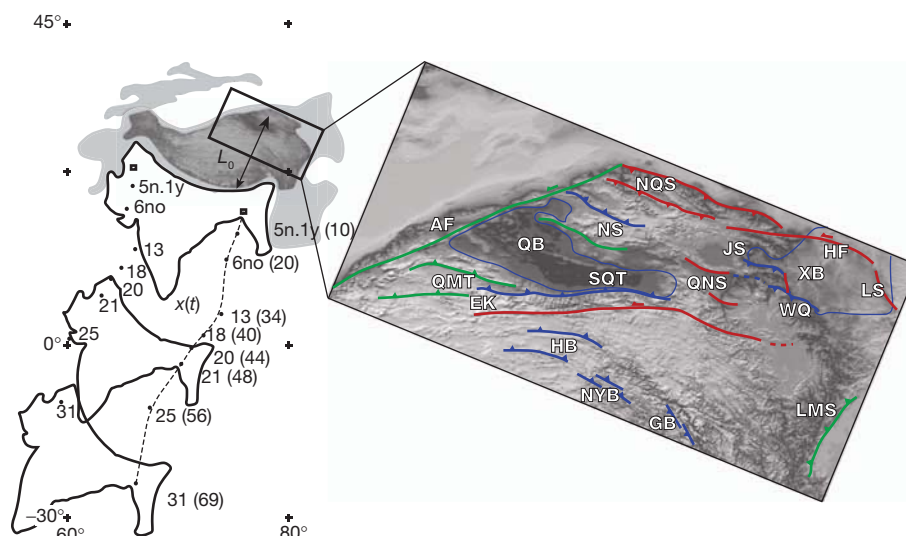


Figure 1 | Plot of Indian plate motion⁷ ($x(t)$) with respect to modern topography of the Tibetan orogen and position of major compressional or transpressional faults. The position of $x(t)$ evaluated here is considered to be the northernmost position of undeformed India such that the region to the north of this point is eventually deformed at some point in the orogen's history and can be included in the determination of $L(t)$. Other arbitrary positions on India would give the same relative path but would include a constant term that would be subtracted from $L(t)$. The map summarizes the initiation of major fault systems (blue, Palaeocene–Eocene; green, Oligocene; red, Miocene–

recent). See Methods for map references. AF, Altyn Tagh; EK, East Kunlun; GB, Gonjo basin; HB, Hoh Xil basin; HF, Haiyuan; JS, Latjie–Jishi Shan; LMS, Longmen Shan; LS, Luopan Shan; NQS, North Qilian Shan; NS, Nan Shan; NYB, Nangqian–Yushu basin; QB, Qaidam basin; QNS, Qinghai Nan Shan; QMT, Qimen Tagh; SQT, South Qaidam; WQ, West Qinling; XB, Xining basin. Numbers next to points identify chron intervals from the geomagnetic timescale and appropriate ages related to reconstructed past positions of two points on the Indian plate with respect to the Eurasia plate⁷. Age (Myr ago) is indicated in parentheses.

deformation of the passive margin of the colliding plate (India), internal deformation to the overriding plate (Eurasia), and southward underthrusting of the Eurasian mantle lithosphere. Shortening in the direction of convergence can also be accommodated by eastward extrusion of Eurasian crust or lithosphere or east–west extension.

Bulk strain rate is defined as compressional in the direction of plate convergence, $\dot{\epsilon} = v/L(t)$, where v is convergence rate (dx/dt) and $L(t)$ is the width of the orogenic belt. The current strain rate is equal to the modern plate convergence rate multiplied by L_0 :

$$dx/dt = \dot{\epsilon} (x(t) + L_0) \quad (2)$$

Hence I consider only the one-dimensional bulk behaviour and do not explicitly consider how convergence is partitioned between thickening and lateral motion or how strain is distributed throughout the plateau.

Next, suppose that $\dot{\epsilon}$ is a constant. Equation (2) would be satisfied by the function

$$x(t) = L_0(e^{-\dot{\epsilon}t} - 1) \quad (3)$$

and would predict that the width of the deforming region and convergence rate should decay exponentially in time. Simply put, if the bulk strain rate is constant, a geometric consequence of a stationary boundary is that plate motion should slow exponentially because the distance that India travels and the rate of motion must be proportional.

Using a nonlinear regression with 95% confidence intervals to the observed past positions of India since 56 Myr ago (ref. 7), I fit equation (3) for two constants L_0 and $\dot{\epsilon}$. Best-fit values of $L_0 = 2,025 \pm 225$ km and $1,174 \pm 93$ km; and $\dot{\epsilon} = 0.0206 \pm 0.001$ and 0.0245 ± 0.001 Myr^{−1} fit points in northeast and northwest India, respectively (Fig. 2 and Supplementary Information). Fitting east and west position data simultaneously produces a best-fit average value of $L_0 = 1,586 \pm 590$ km and $\dot{\epsilon} = 0.0222 \pm 0.005$ Myr^{−1} ($7.03 \times 10^{-16} \pm 1.58 \times 10^{-16}$ s^{−1}) (more details about the statistical fits and differences between the east and west profiles are described in Methods and Supplementary Figs 1 and 2). I compare these values with the modern strain rate in two different ways: first, using the observed length scale or approximate width of

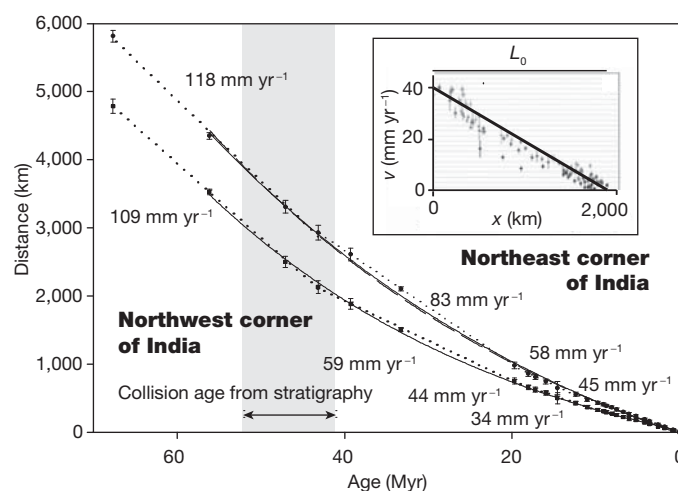


Figure 2 | Predicted versus observed positions of India with respect to present position. The two curves represent reconstructed positions with uncertainties of the northeast and northwest corners of the undeformed Indian plate⁷ from plate circuit models shown in Fig. 1 compared with predicted motion from periods of constant velocity⁷ (dotted lines and velocity values), best-fit nonlinear regression to equation (3) for strain rate and L_0 (solid lines), and best-fit strain rate to equation (3) in which L_0 is fixed to a value of 1,900 km for the eastern promontory (dashed line). Error bars represent 2σ standard errors on position data; 95% confidence and prediction intervals to best-fit functions are shown in Supplementary Figs 1 and 2. Inset: N20°E directed geodetic velocity measurements²⁶ (points) for central and eastern Tibet as a function of distance from the plate boundary compared with a bulk strain rate derived from an observed L_0 value and average plate convergence velocity (solid black line): $\dot{\epsilon} = 6.7 \times 10^{-16}$ s^{−1}. Note that the strain rate derived from plate motions does not specifically require a linear decrease in velocity, only that velocities become negligible at distances greater than L_0 . Deviations of the geodetic velocities from a simple linear trend are undoubtedly due to effects such as the elastic strain accumulation on faults or block rotations, but only the distance over which these velocities decrease rather than the pattern of decrease is of relevance here.

modern Tibet, which when divided by the modern observed convergence rate equals a strain rate; and second, the geodetic contractional strain rate measured over small regions within Tibet. An average distance of about 1,900 km north of the plate boundary coincides with the region where present convergent deformation rates from the Global Positioning System (GPS) in central and northeastern Tibet decrease to within an error of 0 mm yr^{-1} (ref. 26), topography decreases from $\sim 5 \text{ km}$ to $< 1 \text{ km}$, and where mapped active structures become scarce to absent. This distance includes the Qaidam basin and the deforming region of the Qilian Shan to the north. The average value of $L_0 = 1,900$ for an average convergence rate of 40 mm yr^{-1} yields a strain rate of $6.7 \times 10^{-16} \text{ s}^{-1}$ (Fig. 2 inset).

The best-fit value to plate motions suggests a time-averaged bulk shortening strain rate of $(5.45\text{--}8.61) \times 10^{-16} \text{ s}^{-1}$ within 95% confidence. With the exception of the southern Karakorum–Jiali fault zone, interseismic geodetic velocities suggest that the average contractional strain and rotation rates are similar across many sub-regions in Tibet, although geodetic data are more concentrated in the east ($5.23 \times 10^{-16} \text{ s}^{-1}$ to $6.56 \times 10^{-16} \text{ s}^{-1}$ (refs 27, 28) and $4.12 \pm 1.27 \times 10^{-16} \text{ s}^{-1}$ to $5.7 \pm 1.27 \times 10^{-16} \text{ s}^{-1}$ (ref. 26)) in comparison with the $(5.17 \text{ and } 6.7) \times 10^{-16} \text{ s}^{-1}$ values determined from the broadly distributed, linear velocity trend across the plateau in the N32°E and N20°E directions, respectively^{26,27,28} (Fig. 2 inset²⁶). The agreement between long-term and short-term records of contractional strain rate suggests that both the interseismic strain and the time-space-averaged finite strain history from plate motions reflect the same lithospheric property.

For linear or nonlinear viscous rheologies, a constant bulk strain rate implies constant average stress (or constant force) despite the crustal thickening and rise of the Tibetan plateau since collision. I posit that the continental mantle lithosphere creates a constant viscous resistance to plate motion irrespective of crustal buoyancy. This implies either that the positive buoyancy due to crustal thickening is offset by the negative buoyancy of thickening the mantle lithosphere, or that viscous forces dominate over buoyancy forces because the lithosphere is relatively strong²⁹. Small changes in buoyancy could be achieved by having the positive change in GPE created by thickening a buoyant crust offset by thickening dense mantle lithosphere so that the net change in GPE during plateau formation is small ($< 10^{12} \text{ N m}^{-1}$; that is, less than the small change in ridge push force since collision, which is a small positive change in driving force that can be quantified). This condition can be satisfied if the initial Eurasian crust is thin ($< 27 \text{ km}$ thick) for a broad range of reasonable lithospheric density profiles (Fig. 3a).

Alternatively, the constant and low value of strain rate represents the deformation of a homogenous layer within the lithosphere with high viscosity, probably either the continental mantle lithosphere or subhorizontal shear on the Moho. Temperature sensitivities of shear wave velocities have recently been used¹¹ to interpret a region of thick lithosphere beneath Tibet that could represent depleted continental lithosphere attached to the continental crust. Such a region could explain the origin of the uniform strain rate and supports the notion of a relatively strong lithosphere beneath Tibet. In addition, agreement between the observed bulk strain rate and the characteristic length scale of deformation for a thin viscous sheet suggests that buoyancy forces may be negligible compared with viscous forces ($Ar = 0$)³⁰ (Methods; Fig. 3b).

Many studies have concluded that body forces arising from topography on the continents may contribute to plate motion. However, a constant resisting force to plate motion since collision challenges the notion that development of the high terrain in Tibet has altered post-collisional plate motion. Either changes in GPE have been small (that is, if the Eurasian crust was thin before collision and the mantle lithosphere thickens with the crust) or viscous stresses dominate over buoyancy stresses (that is, Tibet is underlain by a deforming, high-viscosity mantle lithosphere). Ultimately, the post-collisional decrease

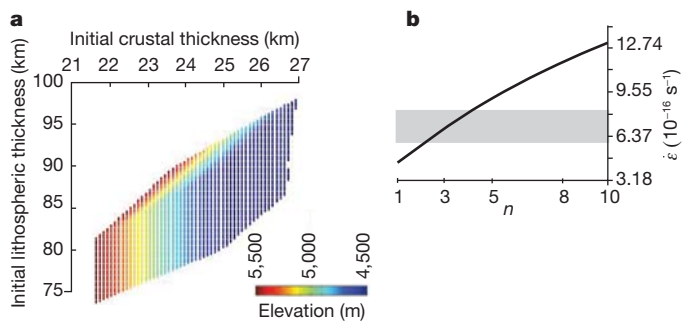


Figure 3 | Models of crust and mantle lithosphere thickening that produce negligible changes in buoyancy and observed bulk strain rates. **a**, Model values of crust and mantle lithosphere thickening that produce neutral buoyancy. Plot of initial lithospheric thickness against initial crustal thickness used to produce an isostatically balanced Tibet-like plateau after crustal thickening to 70 km with a change in GPE of $< 10^{12} \text{ N m}^{-1}$ (ref. 29). Colours represent the final elevation of the lithospheric column after thickening (4,500–5,500 m). The lithospheric density profile is chosen such that the initial lithospheric column has an elevation of 250 m and is in isostatic equilibrium with the mid-ocean ridge column²⁹, using the following limits: $\rho_{\text{co}} = 2,600\text{--}2,800 \text{ kg m}^{-3}$; $\rho_{\text{m}} = 3,150\text{--}3,350 \text{ kg m}^{-3}$; $\alpha = 3.3 \times 10^{-5} \text{ K}^{-1}$. **b**, Estimates of bulk strain rate for thin viscous sheet deformation assuming dominant viscous forces ($Ar = 0$)³⁰. Velocity decays exponentially from the plate boundary as a function of the width of the indenter (2,000 km) and the power-law exponent (n) (curve). Bulk strain rate is determined by the distance from the boundary to the position at which velocities decrease to less than 2.5 mm yr^{-1} ($< 5\%$ of initial value) divided by the boundary velocity (50 mm yr^{-1}) (Methods). The thin viscous sheet model reproduces the observed length scale and thus bulk strain rate for power-law creep exponents $n \approx 2.5\text{--}4$.

in convergence rate comes about not because the plateau grows higher but rather because the orogen narrows under constant forcing and the rate of deformation is controlled by the rheology of the mantle lithosphere. Strength rather than height is a needed departure from the classical interpretation of plate motion for continental margins.

METHODS SUMMARY

I used published positions of northern India since 67 Myr ago (time versus distance)⁷ to evaluate convergence rates between India and Eurasia since continental collision. I determine a best fit to these position data for the equation $x(t) = L_0(e^{-\dot{\epsilon}t} - 1)$ using a nonlinear least-squares regression to observed values with 2σ standard errors. Best-fit values of L_0 and $\dot{\epsilon}$ are interpreted as the modern orogen width and bulk strain rate, respectively. Best-fit values of L_0 were compared with the observed modern width of the plateau determined from topography and GPS, whereas model values of $\dot{\epsilon}$ were compared with observed modern strain rates derived from geodetic data across and within subregions of the plateau. Agreement between long-term strain rates and modern strain rates is used to argue for constant forcing applied to the plateau during post-collisional convergence.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 January 2011; accepted 10 January 2012.

1. Forsyth, D. & Uyeda, S. On the relative importance of the driving forces of plate motion. *Geophys. J. R. Astron. Soc.* **43**, 163–200 (1975).
2. Patriat, P. & Achache, J. India–Eurasia collision chronology has implications for crustal shortening and driving mechanisms of plates. *Nature* **311**, 615–621 (1984).
3. Dewey, J., Cande, S. & Pitman, W. C. Tectonic evolution of the India Eurasia collision zone. *Eclogae Geol. Helv.* **82**, 717–734 (1989).
4. McQuarrie, N., Stock, J. M., Verdel, C. & Wernicke, B. P. Cenozoic evolution of Neotethys and implications for the causes of plate motions. *Geophys. Res. Lett.* **30**, 2026 (2003).
5. Garzione, C., Molnar, P., Libarkin, J. C. & MacFadden, B. J. Rapid late Miocene rise of the Bolivian Altiplano: evidence for removal of the mantle lithosphere. *Earth Planet. Sci. Lett.* **241**, 543–556 (2006).
6. Meade, B. & Conrad, C. Andean growth and the deceleration of South American subduction: time evolution of a coupled orogen-subduction system. *Earth Planet. Sci. Lett.* **275**, 93–101 (2008).
7. Molnar, P. & Stock, J. M. Slowing of India's convergence with Eurasia since 20 Ma and its implications for Tibetan mantle dynamics. *Tectonics* **28**, TC3001 (2009).

8. Copley, A., Avouac, J.-P. & Royer, J.-Y. India–Asia collision and the Cenozoic slowdown of the Indian plate: implications for the forces driving plate motions. *J. Geophys. Res.* **115**, B03410 (2010).
9. Artyushkov, E. V. Stresses in the lithosphere caused by crustal thickness inhomogeneities. *J. Geophys. Res.* **78**, 7675–7708 (1973).
10. Molnar, P. & Lyon-Caen, H. Some simple physical aspects of the support, structure and evolution of mountain belts. *Spec. Pap. Geol. Soc. Am.* **218**, 179–207 (1988).
11. McKenzie, D. & Priestley, K. The influence of lithospheric thickness variations on continental evolution. *Lithos* **102**, 1–11 (2008).
12. Rowley, D. Age of initiation of collision between India and Asia: a review of stratigraphic data. *Earth Planet. Sci. Lett.* **145**, 1–13 (1996).
13. Merkuriev, S. & DeMets, C. Constraints on Indian plate motion since 20 Ma from dense Russian magnetic data: implications for Indian plate dynamics. *Geochem. Geophys. Geosyst.* **7**, Q02002 (2006).
14. Merkuriev, S. & DeMets, C. A high-resolution model for Eurasia–North America plate kinematics since 20 Ma. *Geophys. J. Int.* **173**, 1064–1083 (2008).
15. Flesch, L. M., Haines, A. J. & Holt, W. E. Dynamics of the India–Eurasia collision zone. *J. Geophys. Res.* **106**, 16435–16460 (2001).
16. Kapp, P., DeCelles, P. G., Gehrels, G. E., Heizler, M. & Ding, L. Geological records of the Lhasa–Qiangtang and Indo–Asian collisions in the Nima area of central Tibet. *Geol. Soc. Am. Bull.* **119**, 917–933 (2007).
17. Royden, L., Burchfiel, B. C. & van der Hilst, R. D. The geologic evolution of the Tibetan Plateau. *Science* **321**, 1054–1058 (2008).
18. Clark, M. K., Farley, K. A., Zheng, D. W., Wang, Z. C. & Duvall, A. R. Early Cenozoic faulting of the northern Tibetan Plateau margin from apatite (U–Th)/He ages. *Earth Planet. Sci. Lett.* **296**, 78–88 (2010).
19. Bally, A. W. *Notes on Sedimentary Basins in China. Reports of the American Sedimentary Basins Delegation to the People's Republic of China*. Open File Report USGS 86-327 (1986).
20. Dupont-Nivet, G. *et al.* Paleogene clockwise tectonic rotation of the Xining–Lanzhou region, northeastern Tibetan Plateau. *J. Geophys. Res.* **109**, B04401 (2004).
21. Wang, C. *et al.* Constraints on the early uplift history of the Tibetan Plateau. *Proc. Natl Acad. Sci. USA* **105**, 4987–4992 (2008).
22. Yin, A. *et al.* Tectonic history of the Altyn Tagh fault system in northern Tibet inferred from Cenozoic sedimentation. *Geol. Soc. Am. Bull.* **114**, 1257–1295 (2002).
23. Fang, X., Garzione, C., Van der Voo, R., Li, J. J. & Fan, M. J. Flexural subsidence by 29 Ma on the NE edge of Tibet from the magnetostratigraphy of Linxia Basin, China. *Earth Planet. Sci. Lett.* **210**, 545–560 (2003).
24. Studnicki-Gizbert, C., Burchfiel, B. C., Li, Z. & Chen, Z. Early Tertiary Gonjo basin, eastern Tibet: sedimentary and structural record of the early history of India–Asia collision. *Geosphere* **4**, 713–735 (2008).
25. Spurlin, M. S., Yin, A., Horton, B. K., Zhou, J. & Wang, J. Structural evolution of the Yushu–Nangqian region and its relationship to syncollisional igneous activity, east-central Tibet. *Geol. Soc. Am. Bull.* **117**, 1293–1317 (2005).
26. Zhang, P. *et al.* Continuous deformation of the Tibetan Plateau from global positioning system data. *Geology* **32**, 809–812 (2004).
27. Wang, Q. *et al.* Present-day crustal deformation in China constrained by global positioning system measurements. *Science* **294**, 574–577 (2001).
28. Chen, Q. *et al.* A deforming block model for the present-day tectonics of Tibet. *J. Geophys. Res.* **109**, B01403 (2004).
29. England, P. & Houseman, G. Extension during continental convergence, with application to the Tibetan Plateau. *J. Geophys. Res.* **94**, 17561–17579 (1989).
30. England, P., Houseman, G. & Sonder, L. Length scales for continental deformation in convergent, divergent and strike-slip environments—analytical and approximate solutions for a thin viscous sheet model. *J. Geophys. Res.* **90**, 3551–3557 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements I thank E. Hetland, N. Lewandowski and P. Molnar for discussions and comments, and P. Molnar and J. Stock for sharing their plate reconstruction data in advance of their publication. I also thank Peizhen Zhang and colleagues for leading our collaborative US–Chinese field efforts and cooperative study. L. Flesch and J. Freymueller provided critical feedback that greatly improved the quality of this manuscript. Support for this research was provided by the National Science Foundation, Continental Dynamics Program (EAR-0549748 and EAR-0908711).

Author Contributions M.K.C. performed the data analysis, conceived of the proposed model and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K.C. (marinkc@umich.edu).

METHODS

A trade-off exists between L_0 and $\dot{\epsilon}$ and is shown by a plot of regression values for a range of parameter values (Supplementary Fig. 3). Increasing L_0 requires a corresponding decrease in $\dot{\epsilon}$. If L_0 represents a real value of distance from the plate boundary, changes on the order of 500 km produce differences in strain rate of 30%. A best-fit value for L_0 and $\dot{\epsilon}$ was determined from a nonlinear regression using the Marquardt–Levenberg algorithm in the commercial software SigmaPlot (9.0).

Measured uncertainties from nonlinear regression to the east and west positions independently are smaller (<11%) than the uncertainty when they are considered simultaneously (37%), which may reflect small rotations of India or two-dimensional variations in rheology such as the Tarim basin (Fig. 2 and Supplementary Figs 1 and 2). The plateau geometry and geological history vary from west to east, so perhaps it is not surprising that different values are determined when the west and east are considered independently. The predicted orogen width values based on plate motions are larger than the width of the high, contiguous plateau terrain alone. Therefore the interpretation of orogen width as a measure of bulk shortening length scale requires some inclusion of the less deformed regions such as the Tarim and Qaidam basins and concentrated shortening north of these basins in the Tien Shan and Qilian Shan, respectively. I speculate that the higher value of average strain rate in the west may represent a different average rheology due to the presence of the Tarim Basin in comparison with the plateau in the east. Moreover, L_0 is greater in the east and smaller in the west, and qualitatively the modern width of the plateau matches this asymmetry.

For western Tibet, $L_0 \approx 1,100$ km includes the distance from the Himalayan range front to the northern front of the Tien Shan. Although Oligocene faulting has been documented in the Tien Shan, it is uncertain as to whether Eocene-age faulting occurred in this region. In far eastern Tibet, $L_0 \approx 2,000$ km includes the region north of the plateau that is undergoing eastward motion by extension including the Shanxi rift. L_0 and strain rate trade off with smaller values of L_0 , producing higher strain rates. Forcing the eastern data to $L_0 = 1,900$ km, which is roughly the northern limit of active thrust faulting and thickened crust in northeastern Tibet, produces a best-fit $\dot{\epsilon} = 0.0214 \pm 0.0014 \text{ Myr}^{-1}$, which is 4% higher than when L_0 is not prescribed (Fig. 2). It is interesting to note that inclusion of one point before collision (67 Myr ago) produces slightly higher values for L_0 (10–14%) and lower values for $\dot{\epsilon}$ (7–10%), although exclusion of the point at 56 Myr ago produces negligible change in the regressed values (<1%).

A strong lithosphere is supported by comparing the characteristic length scale of deformation for a thin viscous sheet in which buoyancy forces are negligible in comparison with viscous forces ($Ar = 0$)³⁰ with the observed bulk-strain rate values from plate motions (Fig. 3b). Velocity decays exponentially away from the plate boundary with a characteristic length scale dependent on the width of the indenter (2,000 km for India) and the relationship between stress and strain rate. The speed of the indenter divided by this length scale (L) is an estimate of bulk strain rate. L is taken as the point at which velocities drop to an arbitrarily low value, where I use <5% of the initial 50 mm yr^{-1} average speed of the Indian indenter, which is roughly equal to the uncertainty in the GPS velocities used here.

Where stress and strain rate are related by a power-law function with $n \approx 2.5$ –4, strain rate values from a thin-sheet approximation agree with observed strain rates from plate motions. Values within this range agree with commonly cited n values for power-law creep in olivine. Other stress or strain-rate dependent rheologies cannot be excluded, but would suggest a particular coincidence in changing stress and strain rate during orogen development as to appear constant in time. Low values of the Argand number, Ar , for Tibet do not agree with previous values that suggest that buoyancy forces significantly contribute to the stress field in and around Tibet based on the finite scaling of the plateau³⁰, comparison with modern geodetic and Quaternary fault slip data¹⁵, or shear wave splitting data³¹. However, the low Ar may be compatible with seismic evidence for an attached, depleted lithosphere beneath the plateau¹¹.

The text of the legend to Fig. 1 is repeated here with the detailed references included. Figure 1 | Plot of Indian plate motion ($x(t)$) with respect to modern topography of the Tibet orogen⁷. The map shows different generations of faulting since approximately collision time (blue, Palaeocene–Eocene; green, Oligocene; red, Miocene–recent). Original data references for fault timing are given in the data repository. AF, Altyn Tagh fault³²; EK, East Kunlun fault³³; GB, Gonjo basin²⁴; HB, Hoh Xil basin²¹; HF, Haiyuan fault³⁴; JS, Latjie-Jishi Shan²⁰; LMS, Longmen Shan³⁵; LS, Luipian Shan thrust³⁶; NQS, North Qilian Shan fault³⁷; NS, Nan Shan²²; NYB, Nangqian–Yushu basin²⁵; QB, Qaidam basin¹⁹; QNS, Qinghai Nan Shan³⁷; QMT, Qimen Tagh³⁸; SQT, South Qaidam thrust¹⁸; WQ, West Qinling fault^{18,23,39}; XB, Xining basin²⁰.

1. Flesch, L. M. *et al.* 2005. Constraining the extent of crust–mantle coupling in central Asia using GPS, geologic and shear wave splitting data. *Earth Planet. Sci. Lett.* **238**, 248–268 (2005).
2. Ritts, B. D., Yue, Y. & Graham, S. A. Oligocene–Miocene tectonics and sedimentation along the Altyn Tagh Fault, northern Tibetan Plateau: analysis of the Xorkol, Subei and Aksay basins. *J. Geol.* **112**, 207–229 (2004).
3. Jolivet, M. *et al.* Mesozoic and Cenozoic tectonics of the northern edge of the Tibetan plateau: fission-track constraints. *Tectonophysics* **343**, 111–134 (2001).
4. Burchfiel, B. C. *et al.* Geology of the Haiyuan fault zone, Ningxia–Hui Autonomous region, China, and its relation to the evolution of the northeastern margin of the Tibetan Plateau. *Tectonics* **10**, 1091–1110 (1991).
5. Kirby, E. *et al.* Late Cenozoic evolution of the eastern margin of the Tibetan Plateau: inferences from Ar^{40}/Ar^{39} and (U–Th)/He thermochronometry. *Tectonics* **21**, 1001 (2002).
6. Zheng, D. *et al.* Rapid exhumation at ~8 Ma on the Liupan Shan thrust fault from apatite thermochronology: implications for growth of the northeastern Tibetan Plateau margin. *Earth Planet. Sci. Lett.* **248**, 198–208 (2006).
7. Meyer, B. *et al.* Crustal thickening in Gansu–Qinghai, lithospheric mantle subduction, and oblique, strike-slip controlled growth of the Tibet plateau. *Geophys. J. Int.* **135**, 1–47 (1998).
8. Yin, A., Dang, Y.-Q., Zhang, M., Chen, X.-H. & McRivette, M. W. Cenozoic tectonic evolution of the Qaidam basin and its surrounding regions (Part 3): Structural geology, sedimentation, and regional tectonic reconstruction. *Geol. Soc. Am. Bull.* **120**, 847–876 (2008).
9. Duvall, A. R., Clark, M. K., van der Pluijm, B. A. & Li, C. Direct dating of Eocene reverse faulting in northeastern Tibet using Ar–dating of fault clays and low-temperature thermochronometry. *Earth Planet. Sci. Lett.* **304**, 520–526 (2011).

Circadian rhythms govern cardiac repolarization and arrhythmogenesis

Darwin Jeyaraj^{1,2}, Saptarsi M. Haldar¹, Xiaoping Wan², Mark D. McCauley³, Jürgen A. Ripperger⁴, Kun Hu⁵, Yuan Lu¹, Betty L. Eapen¹, Nikunj Sharma¹, Eckhard Ficker², Michael J. Cutler², James Gulick⁶, Atsushi Sanbe⁶, Jeffrey Robbins⁶, Sophie Demolombe⁷, Roman V. Kondratov⁸, Steven A. Shea⁵, Urs Albrecht⁴, Xander H. T. Wehrens³, David S. Rosenbaum² & Mukesh K. Jain¹

Sudden cardiac death exhibits diurnal variation in both acquired and hereditary forms of heart disease^{1,2}, but the molecular basis of this variation is unknown. A common mechanism that underlies susceptibility to ventricular arrhythmias is abnormalities in the duration (for example, short or long QT syndromes and heart failure)^{3–5} or pattern (for example, Brugada's syndrome)⁶ of myocardial repolarization. Here we provide molecular evidence that links circadian rhythms to vulnerability in ventricular arrhythmias in mice. Specifically, we show that cardiac ion-channel expression and QT-interval duration (an index of myocardial repolarization) exhibit endogenous circadian rhythmicity under the control of a clock-dependent oscillator, krüppel-like factor 15 (*Klf15*). *Klf15* transcriptionally controls rhythmic expression of Kv channel-interacting protein 2 (KChIP2), a critical subunit required for generating the transient outward potassium current⁷. Deficiency or excess of *Klf15* causes loss of rhythmic QT variation, abnormal repolarization and enhanced susceptibility to ventricular arrhythmias. These findings identify circadian transcription of ion channels as a mechanism for cardiac arrhythmogenesis.

Sudden cardiac death from ventricular arrhythmias is the principal cause of mortality from heart disease worldwide and remains a major unresolved public health problem. The incidence of sudden cardiac death exhibits diurnal variation in both acquired and hereditary forms of heart disease^{1,2}. In the general population, the occurrence of sudden cardiac death increases sharply within a few hours of rising in the morning, and a second peak is evident in the evening hours¹. In specific hereditary disorders, for example, Brugada's syndrome, fatal ventricular arrhythmias often occur during sleep². A common mechanism in both acquired and hereditary forms of heart disease that enhances susceptibility to ventricular arrhythmias is abnormal myocardial repolarization⁶. Clinically, three common types of alterations in myocardial repolarization are evident on the surface electrocardiogram (ECG). First, prolongation of repolarization is seen in acquired disorders (for example, heart failure)⁵ and congenital disorders (for example, long QT syndrome)³. Second, shortening of repolarization is found in the short QT syndrome⁴. Third, early repolarization is the hallmark ECG finding in Brugada's syndrome⁸. Interestingly, all three modifications of repolarization increase vulnerability to ventricular arrhythmias⁶. Despite rigorous investigation of the biophysical and structural characteristics of ion channels that control myocardial repolarization, the molecular basis for the diurnal variation in occurrence of ventricular arrhythmias remains unknown.

Biological processes in living organisms that oscillate with a periodicity of 24 h are said to be circadian. This cell-autonomous rhythm is coordinated by an endless negative transcriptional–translational feedback

loop, commonly referred to as the biological clock⁹. Several physiological parameters in the cardiovascular system such as heart rate, blood pressure, vascular tone, QT interval and ventricular effective refractory period exhibit diurnal variation^{10–13}. Recent studies have also identified a direct role for the biological clock in regulating cardiac metabolism, growth and response to injury¹⁴. Previous studies have also reported that expression of repolarizing ion channels and ionic currents (I_{to}) exhibit diurnal changes¹⁵. However, a potential link between circadian rhythms and arrhythmogenesis remains unknown. We made the serendipitous observation that *Klf15* expression exhibits endogenous circadian rhythmicity in the heart (Fig. 1a). Gene expression microarrays in hearts of mice that are deficient in *Klf15* led us to identify *KChIP2* (also called *KCNIP2*), the regulatory β -subunit for the repolarizing transient outward potassium current (I_{to}) as a putative target for this factor in the heart. These observations led us to question whether the circadian clock may regulate rhythmic variation in repolarization and alter susceptibility to arrhythmias through *Klf15*.

First, we explored mechanisms through which the circadian clock regulated rhythmic expression of *Klf15* in the heart. Examination of approximately 5 kb of the promoter region of *Klf15* revealed four canonical 'E-box' regions, that is, consensus binding sites for CLOCK and its heterodimer BMAL1 (also called ARNTL), which are essential transcription factors involved in the circadian clock (Supplementary Fig. 1a, inset). Consistent with this finding, *Klf15* luciferase (approximately 5 kb) was activated in a dose-dependent manner by the CLOCK–BMAL1 heterodimer (Supplementary Fig. 1a). To confirm this interaction, we performed chromatin immunoprecipitation (ChIP) and identified rhythmic variation in BMAL1 binding to the *Klf15* promoter in the hearts of wild-type mice, but not in the hearts of *BMAL1*-null mice (Fig. 1b). In accordance with the observations above, the expression of *Klf15* was disrupted in *Bmal1*-null, and *Per2*- and *Cry1*-null hearts (Supplementary Fig. 1b). Thus, our data strongly suggest that the circadian clock directly regulates the oscillation of *Klf15* in the heart.

To determine whether myocardial repolarization and ion-channel expression exhibit 'true' (endogenous) circadian rhythms—that is, oscillate in the absence of external cues such as light—wild-type mice were placed in constant darkness for 36 h and telemetry-based ECG intervals were measured every 2 h for 24 h. Under these conditions, the heart rate and the QT interval corrected to heart rate (QTc) were both rhythmic and exhibited true endogenous circadian rhythmicity (Fig. 1c, d). Next, to examine whether expression of repolarizing ion channels had endogenous circadian rhythms, mice were placed in constant darkness for 36 h, and hearts were collected every 4 h over a 24-h period. The expression of the α -subunit for the transient outward

¹Case Cardiovascular Research Institute, Harrington Heart and Vascular Institute, Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106, USA. ²Heart and Vascular Research Center, MetroHealth campus of Case Western Reserve University, Cleveland, Ohio 44109, USA. ³Departments of Medicine and Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA. ⁴Department of Medicine, Division of Biochemistry, University of Fribourg, CH-1700 Fribourg, Switzerland. ⁵Division of Sleep Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. ⁶Department of Pediatrics, Division of Molecular Cardiovascular Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ⁷Institut de Pharmacologie Moléculaire et Cellulaire, UMR CNRS 6097, Université de Nice Sophia Antipolis, 06560 Valbonne, France. ⁸Department of Biological, Geological and Environmental Sciences, and Center for Gene Regulation in Health and Disease, Cleveland State University, Cleveland, Ohio, 44115, USA.

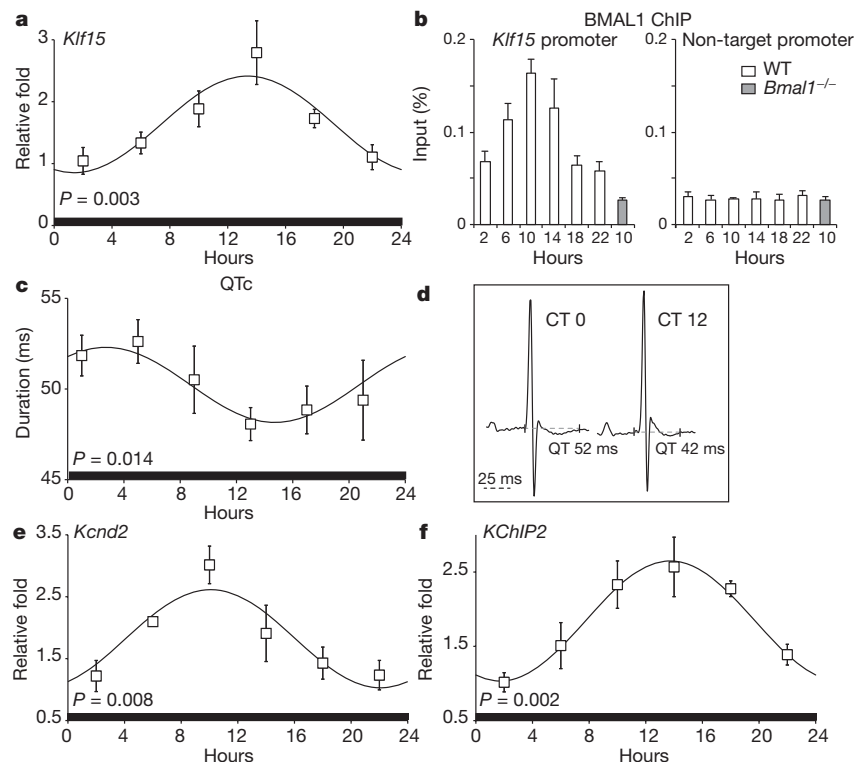


Figure 1 | *Klf15* expression, ECG QTc interval and expression of repolarizing ion channels exhibit endogenous circadian rhythm. **a**, *Klf15* expression exhibits endogenous circadian variation in wild-type (WT) hearts from mice in constant darkness ($n = 4$ per time point). CT, circadian time. **b**, Effect of BMAL1 ChIP on the *Klf15* promoter, showing rhythmic variation in binding of BMAL1 to the *Klf15* promoter in wild-type hearts ($n = 3$ per group). **c**, Duration of ECG QTc interval (ms) in conscious mice exhibits endogenous circadian variation in constant darkness ($n = 4$). **d**, Representative ECGs from conscious mice after 36 h in constant darkness at CT 0 and CT 12. **e**, **f**, Endogenous circadian variation in transcripts for *Kcnd2* and *KChIP2* in wild-type hearts measured every 4 h after 36 h in constant darkness ($n = 4$ per time point). Error bars, mean \pm s.e.m.

potassium current (I_{to}), Kv4.2 (encoded by *Kcnd2*) (Fig. 1e), and the regulatory β -subunit, *KChIP2* (Fig. 1f), exhibit endogenous circadian rhythmicity, as did components of the circadian clock in the heart (Supplementary Fig. 2). In contrast, the expression of two other major repolarizing currents in the murine ventricle, Kv1.5 (the α -subunit for the ultra-rapid delayed rectifier potassium current) and Kir2.1 (the α -subunit for the inward rectifier potassium current), did not reveal notable rhythmic variation (Supplementary Fig. 3). In addition, we observed a 24-h rhythm in the oscillation of *Bmal1*, *Klf15* and *KChIP2* after serum shock in cultured neonatal rat ventricular myocytes (Supplementary Fig. 4). These data indicate that myocardial repolarization and the expression of some repolarizing ion channels exhibit an endogenous circadian rhythm.

Next, to elucidate the role of *Klf15* in regulating rhythmic changes in repolarization, we used complementary *in vivo* loss- and gain-of-function approaches in mice. For loss-of-function, a previously described systemic *Klf15*-null mouse was used¹⁶; for gain-of-function, a cardiac-specific *Klf15* transgenic (*Klf15*-Tg) mouse driven by an attenuated α -myosin heavy chain (α -MHC) promoter was developed (Supplementary Fig. 5). First, we examined whether rhythmic expression of *Kcnd2* or *KChIP2* was altered in the *Klf15*-deficient state. *Kcnd2* expression exhibited altered rhythmic variation in *Klf15*-null mice with reduced expression at zeitgeber time 6 (ZT6), and increased expression at ZT22 compared to wild-type controls (Fig. 2a). *KChIP2* expression was devoid of any discernable rhythm in the *Klf15*-null mice and sustained reduction was observed at all time points (Fig. 2b, c and

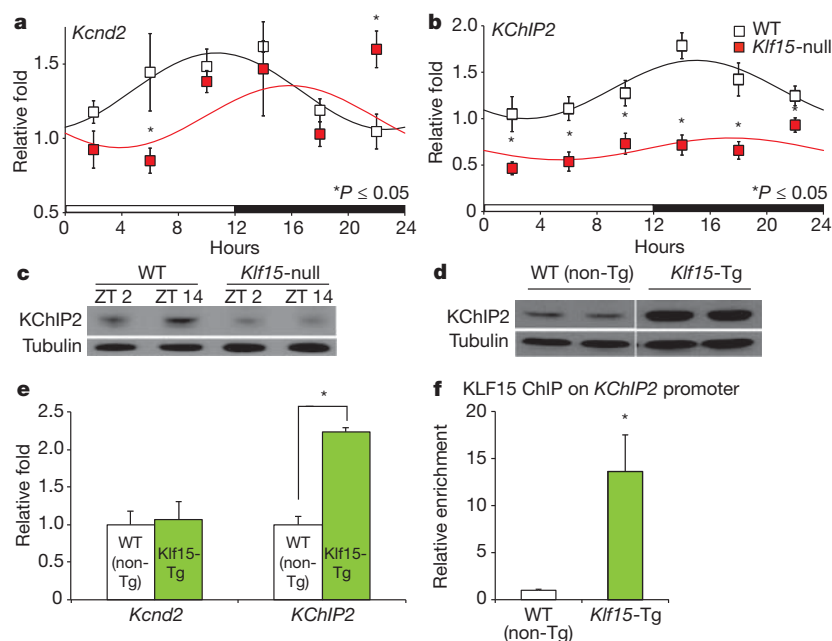


Figure 2 | *Klf15* regulates *KChIP2* expression in the heart. **a**, *Kcnd2* mRNA expression exhibits diurnal rhythm in wild-type mice ($P = 0.0023$), but in *Klf15*-null hearts (P not significant) the rhythm is abnormal with reduced expression at zeitgeber time 6 (ZT6) and increased expression at ZT22 ($n = 4$ per time point per group). **b**, *KChIP2* mRNA expression exhibits no rhythmic variation in *Klf15*-deficient mice (WT, $P = 0.016$; *Klf15*-null, P not significant), with substantial reductions in expression at all time points ($n = 4$ per time point per group). **c**, *KChIP2* protein expression exhibits no variation over 12 h in *Klf15*-null hearts. **d**, **e**, *Klf15*-Tg mice hearts express higher levels of *KChIP2* mRNA and protein. **f**, Chromatin immunoprecipitation with Flag antibody illustrating enrichment of Flag-*KLF15* on the *KChIP2* promoter ($n = 3$ per group). Error bars, mean \pm s.e.m., $*P < 0.05$.

Supplementary Fig. 6a). Next, we examined whether *Kcnd2* or *KChIP2* serve as transcriptional targets for *Klf15* in the heart. Adenoviral over-expression of *Klf15* in neonatal rat ventricular myocytes robustly induced *KChIP2* expression but had no effect on *Kcnd2* expression (Supplementary Fig. 6b). Notably, in *Klf15*-Tg hearts, expression of *KChIP2* was twofold greater but with no effect on *Kcnd2* expression (Fig. 2d, e). Examination of the *KChIP2* promoter region revealed numerous consensus krüppel-binding sites, that is, C(A/T)CCC (Supplementary Fig. 7a). The activity of *KChIP2* luciferase was induced by full-length KLF15 but not by a mutant that lacked the zinc-finger DNA-binding domain (Supplementary Fig. 7b). To identify the specific *Klf15* binding site, deletion constructs of the *KChIP2* promoter were generated, and transcriptional activity was mapped to the proximal 555 bases (Supplementary Fig. 7a). Mutation of one krüppel-binding site within this region ($\Delta 1$) was sufficient to cause complete loss of activity in the full-length *KChIP2* promoter (Supplementary Fig. 7c). Chromatin immunoprecipitation of Flag-KLF15 from *Klf15*-Tg hearts confirmed that KLF15 was enriched on the endogenous *KChIP2* promoter (Fig. 2f). Importantly, the oscillation of several components of the core clock machinery was minimally affected in the *Klf15*-deficient state (Supplementary Fig. 8). In addition, the expression levels of clock genes in *Klf15*-Tg hearts were similar to their controls at ZT6 (Supplementary Fig. 8). This suggested that the endogenous clock is dependent on *Klf15* to orchestrate rhythmic changes in *KChIP2* expression. Consistent with this observation, the expression of *Klf15* (Supplementary Fig. 1b) and *KChIP2* (Supplementary Fig. 9) were altered in a similar fashion in *Bmal1*-null, and *Per2*- and *Cry1*-null mice. These data support the idea that *KChIP2* is a direct transcriptional target for *Klf15* in the heart.

We next examined whether *Klf15*-dependent regulation of *KChIP2* could be responsible for rhythmic day–night variation in myocardial repolarization. Analysis of telemetry-based ECGs revealed that rhythmic QTc interval variation was indeed abrogated in both *Klf15*-null and *Klf15*-Tg mice (Fig. 3a–d). In the *Klf15*-deficient state, the ECG QTc interval was prolonged in the dark phase and failed to oscillate (Fig. 3a, c). This occurred despite *Klf15*-null mice having similar heart rates to their wild-type counterparts (Supplementary Fig. 11). In contrast, the *Klf15*-Tg mice had persistently short QT intervals with no rhythmic day–night variation (Fig. 3b, d). Again, this occurred despite minimal difference in heart rates when compared to wild-type controls (Supplementary Fig. 11). Next, we examined whether transient outward current ($I_{to\ fast}$)-dependent changes in repolarization in isolated myocytes were responsible for the ECG changes mentioned above in *Klf15*-null and *Klf15*-Tg mice. In *Klf15*-null mice, there was a marked reduction in $I_{to\ fast}$ density (Fig. 3e) and prolongation of action potential duration (APD) (Fig. 3g). In contrast, *Klf15*-Tg mice exhibited a substantial increase in $I_{to\ fast}$ density (Fig. 3f) with a dramatic shortening of APD (Fig. 3h). In the *Klf15*-Tg mice, in addition to short QT intervals, we observed ST-segment changes indicative of early repolarization that are similar to ECG findings in Brugada's syndrome⁸ (Fig. 3b, arrows). Our data suggest that *Klf15*-dependent transcriptional regulation of rhythmic *KChIP2* expression in murine hearts plays a central part in rhythmic variation in ventricular repolarization.

Next, we examined whether excessive prolongation or shortening of repolarization could alter arrhythmia susceptibility and survival. *Klf15*-null mice show no spontaneous arrhythmias on ECG telemetry, hence we used intracardiac programmed electrical stimulation to examine arrhythmia susceptibility. In contrast to wild-type mice, a marked increase in occurrence of ventricular arrhythmias was seen in *Klf15*-null mice (Fig. 4a). Notably, *Klf15*-Tg mice exhibit spontaneous ventricular arrhythmias on ECG telemetry (Fig. 4b) and succumb to ~35% mortality by 4 months of age (three out of eight deaths in *Klf15*-Tg versus no deaths out of eight in wild-type non-transgenic controls, data not shown). As the *Klf15*-null mice show no evidence of overt ventricular dysfunction, apoptosis or fibrosis^{16,17} in the basal state, the enhanced susceptibility to arrhythmias is probably primarily driven by

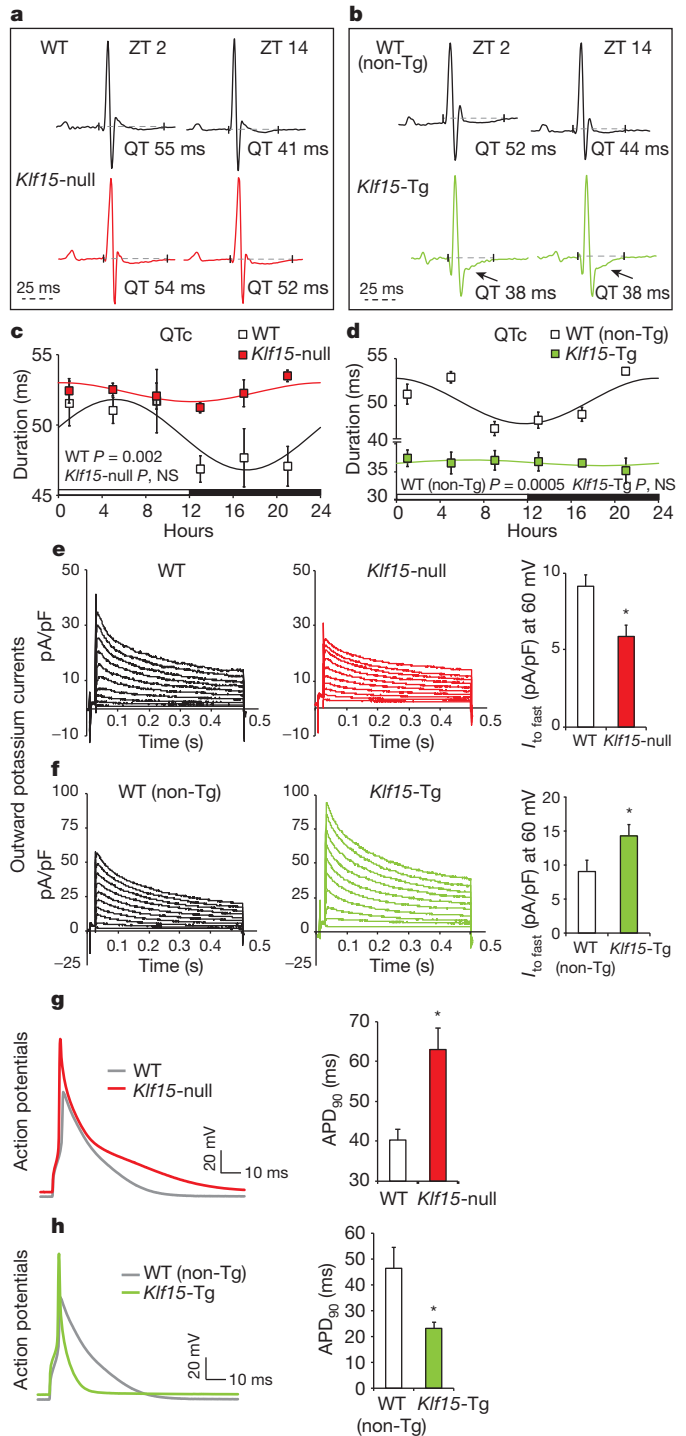


Figure 3 | Deficiency or excess of *Klf15* modulates rhythmic variation in repolarization. **a, b**, Representative ECGs from wild-type versus *Klf15*-null mice, and wild-type (non-Tg) versus *Klf15*-Tg mice at ZT2 and ZT14. Note the ST-segment abnormalities in *Klf15*-Tg mice (arrows). **c**, QTc interval exhibits 24-h rhythm in wild-type mice; this rhythm is abrogated with prolonged QTc in the dark phase in *Klf15*-null mice ($n = 4$ for wild type, $n = 4$ for *Klf15*-null). **d**, *Klf15*-Tg mice exhibit persistently short QT intervals with no day–night rhythmic variation compared to wild-type (non-Tg) controls ($n = 3$ for wild type, $n = 4$ for *Klf15*-Tg). **e, f**, Representative outward current recordings from all study groups and summary data for the amplitude of $I_{to\ fast}$ measured at 60 mV with an average time of decay of 45 ± 5 ms ($n = 10$ for wild type, $n = 13$ for *Klf15*-null; $n = 14$ for wild type (non-Tg), and $n = 19$ for *Klf15*-Tg). **g, h**, Representative ventricular action potentials from all study groups with summary data in bar graphs ($n = 10$ for wild type, $n = 13$ for *Klf15*-null; $n = 14$ for wild type (non-Tg), and $n = 19$ *Klf15*-Tg). Error bars, mean \pm s.e.m., $*P < 0.05$. APD₉₀, action potential duration measured at 90% repolarization.

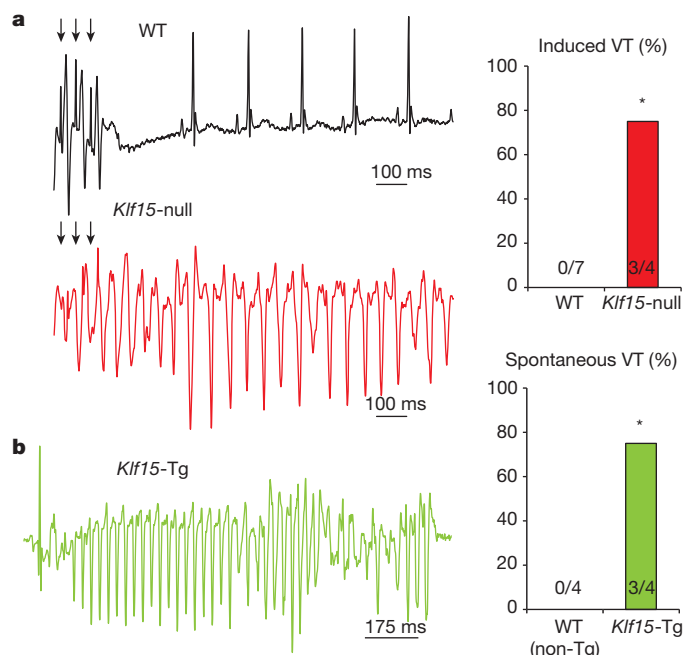


Figure 4 | *Klf15* deficiency or excess increases susceptibility to ventricular arrhythmias. **a**, Programmed electrical stimulation in wild-type and *Klf15*-null mice. Onset of ventricular tachycardia after premature stimuli is shown (arrows) in *Klf15*-null mice (none of the seven wild-type mice were inducible but three of the four *Klf15*-null mice were inducible; * $P < 0.05$). **b**, Spontaneous ventricular arrhythmia in *Klf15*-Tg mice. (none of the four wild-type mice exhibited spontaneous arrhythmias but three of the four *Klf15*-Tg mice exhibited ventricular arrhythmias; * $P < 0.05$). VT, ventricular tachycardia.

abnormalities in repolarization. Our studies demonstrate that both deficiency and excess of *Klf15* impair temporal variation in cardiac repolarization and greatly increase susceptibility to arrhythmias.

Although our finding of circadian control of KChIP2 by *Klf15* establishes the principle that circadian rhythms may contribute to arrhythmogenesis, we note that *Klf15* minimally affects *Kcnd2* expression that also exhibits circadian rhythm (Fig. 1f). However, *Kcnd2* expression was disrupted in *Bmal1*-null and *Per2*- and *Cry1*-null hearts, and this is indicative of a direct regulation by the circadian clock (Supplementary Fig. 12). Consistent with this observation, cardiomyocytes from *Bmal1*-null mice exhibit marked action potential prolongation due to near-complete elimination of the fast component of the transient outward potassium current (Supplementary Fig. 13). This raises the possibility that additional factors—perhaps components of the circadian clock or unidentified transcriptional regulators—may also affect temporal variation in electrophysiological parameters and arrhythmogenesis. Future studies in cardiac-specific deletion of clock components would be necessary to confirm whether the ion channel rhythms are cell autonomous, and their role in regulating cardiac electrophysiology.

Our study provides the first mechanistic link between endogenous circadian rhythms and the cardiac electrical instability that is most often associated with sudden cardiac death in humans (Supplementary Fig. 14). Specifically, we show that *Klf15*-dependent rhythmic transcription of *KChIP2* regulates the duration and pattern of repolarization and susceptibility to arrhythmias in mice. As the occurrence of sudden cardiac death in acquired and hereditary forms of human heart disease follows a distinct diurnal pattern^{1,2}, these observations offer new insights into unrecognized triggers of electrical instability in the heart. However, in contrast to murine repolarization, which is largely dependent on I_{to} , human repolarization occurs through a complex interaction of multiple repolarizing ionic currents. Thus, additional studies will be needed to develop a comprehensive understanding of the link between the circadian clock and electrophysiological properties

of the human heart. Nevertheless, these data may provide a mechanistic foundation for future efforts to prevent or treat cardiac arrhythmias by modulating the circadian clock through behavioural or pharmacological means.

METHODS SUMMARY

Mice used in the present study, messenger RNA quantification using polymerase chain reaction with reverse transcription (RT-PCR), promoter reporter analysis, western immunoblot analysis, chromatin immunoprecipitation, telemetry ECG and interval analysis, isolated myocyte studies for action potential or I_{to} measurements, *in vivo* electrophysiological studies for arrhythmia susceptibility, cosinor analysis for rhythm assessment, and statistical methods are detailed in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 March 2011; accepted 12 January 2012.

Published online 22 February 2012.

- Muller, J. E. *et al.* Circadian variation in the frequency of sudden cardiac death. *Circulation* **75**, 131–138 (1987).
- Matsuo, K. *et al.* The circadian pattern of the development of ventricular fibrillation in patients with Brugada syndrome. *Eur. Heart J.* **20**, 465–470 (1999).
- Goldenberg, I. & Moss, A. J. Long QT syndrome. *J. Am. Coll. Cardiol.* **51**, 2291–2300 (2008).
- Patel, U. & Pavri, B. B. Short QT syndrome: a review. *Cardiol. Rev.* **17**, 300–303 (2009).
- Tomaselli, G. F. & Marban, E. Electrophysiological remodeling in hypertrophy and heart failure. *Cardiovasc. Res.* **42**, 270–283 (1999).
- Antzelevitch, C. Role of spatial dispersion of repolarization in inherited and acquired sudden cardiac death syndromes. *Am. J. Physiol. Heart Circ. Physiol.* **293**, H2024–H2038 (2007).
- Kuo, H. C. *et al.* A defect in the Kv channel-interacting protein 2 (KChIP2) gene leads to a complete loss of I_{to} and confers susceptibility to ventricular tachycardia. *Cell* **107**, 801–813 (2001).
- Antzelevitch, C. & Yan, G. X. J wave syndromes. *Heart Rhythm* **7**, 549–558 (2010).
- Reppert, S. M. & Weaver, D. R. Coordination of circadian timing in mammals. *Nature* **418**, 935–941 (2002).
- Bexton, R. S., Vallin, H. O. & Camm, A. J. Diurnal variation of the QT interval— influence of the autonomic nervous system. *Br. Heart J.* **55**, 253–258 (1986).
- Kong, T. Q. Jr, Goldberger, J. J., Parker, M., Wang, T. & Kadish, A. H. Circadian variation in human ventricular refractoriness. *Circulation* **92**, 1507–1516 (1995).
- Martino, T. A. & Sole, M. J. Molecular time: an often overlooked dimension to cardiovascular disease. *Circ. Res.* **105**, 1047–1061 (2009).
- Paschos, G. K. & FitzGerald, G. A. Circadian clocks and vascular function. *Circ. Res.* **106**, 833–841 (2010).
- Durgan, D. J. & Young, M. E. The cardiomyocyte circadian clock: emerging roles in health and disease. *Circ. Res.* **106**, 647–658 (2010).
- Yamashita, T. *et al.* Circadian variation of cardiac K^+ channel gene expression. *Circulation* **107**, 1917–1922 (2003).
- Haldar, S. M. *et al.* *Klf15* deficiency is a molecular link between heart failure and aortic aneurysm formation. *Sci. Transl. Med.* **2**, 26ra26 (2010).
- Wang, B. *et al.* The Krüppel-like factor KLF15 inhibits connective tissue growth factor (CTGF) expression in cardiac fibroblasts. *J. Mol. Cell. Cardiol.* **45**, 193–197 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. F. Connors Jr for support, M. Mustar for illustrations, Y. Cui for experimental assistance, and members of the Jain laboratory for discussions. Funding sources: Heart Rhythm Society Fellowship (D.J.); National Institutes of Health grants HL094660 (D.J.), HL066991 (M.D.M.), HL086614 (S.M.H.), American Heart Association postdoctoral grant (N.S.), HL089598, HL091947 (X.H.W.), HL76446 (S.A.S.), HL102241 (K.H.), HL054807 (D.S.R.), HL075427, HL076754, HL084154, HL086548 and HL097595 (M.K.J.); Swiss National Science Foundation grants 31003A/131086 (U.A.) and M01-RR02635 (B.W.H.); Leducq Foundation grants of the ENAFRA Network 07CVD03 (S.D.); and the Centre National de la Recherche Scientifique (S.D.).

Author Contributions D.J. and M.K.J. designed the research; D.J., S.M.H., X.W., M.D.M., J.A.R., Y.L., B.L.E. and M.J.C. carried out the experiments; J.G., A.S., J.R. and R.V.K. contributed critical reagents; D.J., N.S., S.D., R.V.K., S.A.S., U.A., X.H.T.W., D.S.R. and M.K.J. supervised the research; D.J., S.M.H., X.W., M.D.M., J.A.R., K.H., B.L.E., E.F., S.A.S., U.A., X.H.T.W., D.S.R. and M.K.J. analysed and interpreted the data; and D.J. and M.K.J. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K.J. (mukesh.jain2@case.edu) or D.J. (darwinjeyaraj@gmail.com).

METHODS

Mice. All animal studies were carried out with permission, and in accordance with, animal care guidelines from the Institutional Animal Care Use Committee (IACUC) at Case Western Reserve University and at collaborating facilities. Wild-type male mice on C57BL6/J background (Jackson Laboratory) were bred in our facility and used for circadian studies. Mice were housed under strict light–dark conditions (lights on at 6:00 and lights off at 18:00) and had free access to standard chow and water, and were minimally disturbed for 4–6 weeks before the final experiment. Generation of systemic *Klf15*-null mice was as described previously¹⁸. *Klf15*-null mice have been backcrossed into the C57BL6/J background for over ten generations¹⁸ and the BMAL1 mice were bred as previously described¹⁹. For *Klf15*-Tg mice, Flag-KLF15 was cloned downstream of an attenuated α -myosin heavy-chain promoter as previously described²⁰. This construct was injected into FVB (friend leukemia virus B mouse strain) oocytes, and after germline transmission the mice were examined for expression of the transgene. Wild-type (non-Tg) littermates served as controls. For light–dark experiments, mice were killed with CO₂ inhalation or isoflurane every 4 h for 24 h. For constant dark experiments, mice were placed in complete darkness for 36 h (starting at the end of light phase at ZT12) and hearts were collected every 4 h over a 24-h period.

RNA isolation and RT-PCR analysis. After euthanasia, hearts were collected, washed in cold phosphate buffered saline, the atria removed and the ventricles dissected to the apical and basal regions, and flash frozen in liquid nitrogen. RNA was isolated from the apical regions of frozen heart samples by homogenization in Trizol reagent (Invitrogen) by following the manufacturer's instructions (Invitrogen). RNA was reverse transcribed after DNase treatment (New England Biolabs). RT-PCR was performed using locked nucleic acid (LNA)-based TaqMan approach with primers and probes designed, and their efficiency tested, at the Universal Probe Library (Roche), and with β -actin used as the normalizing gene.

Cell-culture studies. Neonatal rat ventricular myocytes were isolated from 1–2-day-old rat pups and grown under standard conditions¹⁸. Adenoviral overexpression was performed for 24 h and myocytes were then collected for mRNA and protein analysis. For synchronization, the myocytes were starved in media containing insulin, transferrin and selenium (ITS supplement, Sigma-Aldrich) for 48 h. After this, the myocytes were synchronized with 50% horse serum for 30 min, washed twice with no-serum media and replenished with ITS-containing media. The mouse *Klf15* promoter (approximately 5 kb) was cloned into PGL3-basic (Promega). The rat *KChIP2* luciferase was a gift from P. H. Backx. Mutant constructs of rat *KChIP2* luciferase were generated by PCR-based TOPO cloning (Invitrogen), and site-directed mutagenesis was performed using Quikchange II mutagenesis kit (Agilent Technologies) and confirmed by sequencing. *Klf15* and *KChIP2* luciferase studies were conducted in NIH3T3 cells, and luciferase activity was normalized to protein concentration.

Western immunoblot analysis. For detecting Flag-KLF15, nuclear lysates were prepared using the NE-PER kit following manufacturer's instructions (Thermo Scientific) and probed with anti-Flag antibody (Sigma). For KChIP2 analysis, whole-cell lysates were prepared by homogenizing the basal regions of the hearts in buffer containing Tris-HCl (50 mM, pH 7.4), NaCl (150 mM), NP-40 (1%), sodium deoxycholate (0.25%), EDTA (1 mM), and supplemented with protease and phosphatase inhibitors (Roche). The blots were probed with a mouse monoclonal antibody against KChIP2 (NIH Neuromab), normalized to tubulin (Sigma-Aldrich) and quantified using Quantity One software (Bio-Rad).

ChIP. ChIP was performed with hearts as previously described^{21,22}. In brief, hearts were fixed with fresh 1.11% formaldehyde for 10 min, and then by chromatin preparation and sonication (Diagenode). The sonicated chromatin was immunoprecipitated using BMAL1 or Flag antibody bound to Dynabeads (Invitrogen). The relative abundance was normalized to abundance of 28S between the input and immunoprecipitated samples as previously described²¹. Primers that were used for BMAL1 ChIP on the *Klf15* promoter were; forward, 5'-GCCTGAGCATCTCCCCATCA-3'; reverse, 5'-GGGGCCACCTCTCTGGACTT-3'; and probe, 5'-FAM-CCCGCCAGTGACCATGTCTGCCTGT-3'-BHQ1. Non-target primers were; forward, 5'-GCCAATTCACATTTCAACCA-3'; reverse, 5'-GACACAAGGCATTTCAA-3'; and probe, 5'-FAM-TGCAAAGGGCTGGGACATGGG-3'-BHQ1. Primers that were used for ChIP of Flag-KLF15 on the *KChIP2* promoter were; forward, 5'-GCTCCGCTCTCACTTGCT-3'; and reverse, 5'-GGCTGGCAAGGCTTTTCT-3'.

Telemetry ECG and interval analysis. Mice were implanted with telemetry devices (ETA F20, Data Sciences International) and allowed to recover for at least 2 weeks. ECGs were recorded from conscious mice continuously in their native environment and digital data (PhysioTel, Data Sciences International) were stored for future analysis. Owing to rapid changes in the mouse heart rates, a weighted heart-rate approach was used to assess rhythmic changes in QT interval, and measurements were made every 2 h over a 24-h period. First, the average heart rate was calculated for each hour by digital tracking of the ECG RR intervals (time

interval between two consecutive R waves) using the Dataquest analysis software (Data Sciences International). Then, during the first instance within each hour when the average heart rate was present, the QT interval was measured using electronic calipers from two consecutive beats. The QT interval was corrected for heart rate using a previously validated formula for conscious mice $QT/(RR/100)^{1/2}$ (ref. 23). A Cosinor model was applied to assess the 24-h rhythm in QT using a sinusoidal regression function and raw data presented in four hourly blocks for visualization purposes.

Electrophysiological studies in myocytes. Murine ventricular myocytes were isolated using a standard enzymatic dispersion technique following overnight fast as previously described²⁴. Myocytes were re-suspended in media 199, allowed to recover and recordings were conducted within several hours on the same day. The conventional whole-cell mode was used to record action potentials and I_{to} . In brief, myocytes were bathed in a chamber that was continuously perfused with Tyrode's solution of the following composition (in mmol l⁻¹): NaCl, 137; KCl, 5.4; CaCl₂, 2.0; MgSO₄, 1.0; glucose, 10; and HEPES, 10 (pH 7.35). Patch pipettes (0.9–1.5 M Ω) were filled with electrode solution composed of (in mmol l⁻¹): aspartic acid, 120; KCl, 20; NaCl, 10; MgCl₂, 2; and HEPES, 5 (pH 7.3). Action potentials were elicited in current-clamp mode by injection of a square pulse of current of 5 ms duration and 1.5–2 times the threshold amplitude. APD was measured at 90% repolarization. To measure I_{to} , cells were placed in Tyrode's solution (as described earlier) containing 1 μ M nisoldipine to block calcium current and calcium-activated chloride current, and tetrodotoxin (100 μ mol l⁻¹) to block sodium current. Cells were brought from a holding potential of –70 mV to –25 mV for 25 ms. To isolate the fast, transient component of the outward currents, $I_{to fast}$, the decay phase of outward potassium currents was fit by the exponential functions of the form:

$$y(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2) + A_{ss}$$

where τ_1 is the time constant of decay of the fast, transient component of outward potassium currents; A_1 is the amplitude coefficient of $I_{to fast}$; τ_2 is the time constant of decay of the slow, transient component of the outward currents; A_2 is the amplitude of $I_{to slow}$; and A_{ss} is the amplitude coefficient of the non-inactivating steady-state outward potassium current I_{ss} . Consistent with previous studies²⁵, the time constant of decay of the fast, transient component $I_{to fast}$ was 46 ± 5 ms. The measured current amplitudes were normalized to cell capacitance and converted into current densities. All experiments were conducted at 36 °C. Cell capacitance and series resistance were compensated electronically at ~80%. Command and data acquisition were operated with an Axopatch 200B patch-clamp amplifier controlled by a personal computer using a Digidata 1200 acquisition board driven by pCLAMP 7.0 software (Axon Instruments).

Programmed electrical stimulation. Intracardiac programmed electrical stimulation was performed as previously described²⁶. In brief, mice were anaesthetized using 1.5% isoflurane in 95% O₂ after an overnight fast. ECG channels were amplified (0.1 mV cm⁻¹) and filtered between 0.05 and 400 Hz. A computer-based data acquisition system (Emka Technologies) was used to record a 3-lead body surface ECG, and up to four intracardiac bipolar electrograms. Bipolar right atrial pacing and right ventricular pacing were performed using 2-ms current pulses delivered by an external stimulator (STG-3008, MultiChannel Systems; Reutlingen). Standard clinical electrophysiologic pacing protocols were used to determine all basic electrophysiologic parameters. Overdrive pacing, single, double and triple extrastimuli, as well as ventricular burst pacing, were delivered to determine the inducibility of ventricular arrhythmias, which was tested twice.

Statistical analysis. A cosinor model was adopted to determine whether there is a substantial 24-h rhythm in each physiological and molecular variable of interest. By pooling data points of all mice, the model fits data to a fundamental sinusoidal function²⁷. To determine the coefficients (amplitude and phase) of the sinusoidal function and to see whether there were significant relationships, a mixed model analysis of variance was performed using standard least-square regression and the restricted maximum likelihood method (JMP 8.0, SAS Institute) as previously described²⁸. Data are presented as mean \pm s.e.m., the Student's *t*-test was used for assessing the difference between individual groups and $P \leq 0.05$ was considered statistically significant.

- Fisch, S. *et al.* Krüppel-like factor 15 is a regulator of cardiomyocyte hypertrophy. *Proc. Natl Acad. Sci. USA* **104**, 7074–7079 (2007); correction **104**, 13851 (2007).
- Bunger, M. K. *et al.* Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell* **103**, 1009–1017 (2000).
- Sanbe, A. *et al.* Reengineering inducible cardiac-specific transgenesis with an attenuated myosin heavy chain promoter. *Circ. Res.* **92**, 609–616 (2003).
- Tuteja, G., Jensen, S. T., White, P. & Kaestner, K. H. *Cis*-regulatory modules in the mammalian liver: composition depends on strength of Foxa2 consensus site. *Nucleic Acids Res.* **36**, 4149–4157 (2008).
- Ripperger, J. A. & Schibler, U. Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian *Dbp* transcription and chromatin transitions. *Nature Genet.* **38**, 369–374 (2006).

23. Mitchell, G. F., Jeron, A. & Koren, G. Measurement of heart rate and Q-T interval in the conscious mouse. *Am. J. Physiol.* **274**, H747–H751 (1998).
24. Libbus, I., Wan, X. & Rosenbaum, D. S. Electrotonic load triggers remodeling of repolarizing current I_{to} in ventricle. *Am. J. Physiol. Heart Circ. Physiol.* **286**, H1901–H1909 (2004).
25. Wagner, S. *et al.* Ca/calmodulin kinase II differentially modulates potassium currents. *Circ. Arrhythm. Electrophysiol.* **2**, 285–294 (2009).
26. van Oort, R. J. *et al.* Ryanodine receptor phosphorylation by calcium/calmodulin-dependent protein kinase II promotes life-threatening ventricular arrhythmias in mice with heart failure. *Circulation* **122**, 2669–2679 (2010).
27. Nelson, W., Tong, Y. L., Lee, J. K. & Halberg, F. Methods for cosinor-rhythmometry. *Chronobiologia* **6**, 305–323 (1979).
28. Hu, K., Scheer, F. A., Laker, M., Smales, C. & Shea, S. A. Endogenous circadian rhythm in vasovagal response to head-up tilt. *Circulation* **123**, 961–970 (2011).

Structural basis for recognition of H3K56-acetylated histone H3–H4 by the chaperone Rtt106

Dan Su¹*, Qi Hu¹*, Qing Li¹*, James R. Thompson², Gaofeng Cui¹, Ahmed Fazly¹, Brian A. Davies¹, Maria Victoria Botuyan¹, Zhiguo Zhang¹ & Georges Mer¹

Dynamic variations in the structure of chromatin influence virtually all DNA-related processes in eukaryotes and are controlled in part by post-translational modifications of histones^{1–3}. One such modification, the acetylation of lysine 56 (H3K56ac) in the amino-terminal α -helix (α N) of histone H3, has been implicated in the regulation of nucleosome assembly during DNA replication and repair, and nucleosome disassembly during gene transcription^{4–10}. In *Saccharomyces cerevisiae*, the histone chaperone Rtt106 contributes to the deposition of newly synthesized H3K56ac-carrying H3–H4 complex on replicating DNA⁵, but it is unclear how Rtt106 binds H3–H4 and specifically recognizes H3K56ac as there is no apparent acetylated lysine reader domain in Rtt106. Here, we show that two domains of Rtt106 are involved in a combinatorial recognition of H3–H4. An N-terminal domain homodimerizes and interacts with H3–H4 independently of acetylation while a double pleckstrin-homology (PH) domain binds the K56-containing region of H3. Affinity is markedly enhanced upon acetylation of K56, an effect that is probably due to increased conformational entropy of the α N helix of H3. Our data support a mode of interaction where the N-terminal homodimeric domain of Rtt106 intercalates between the two H3–H4 components of the (H3–H4)₂ tetramer while two double PH domains in the Rtt106 dimer interact with each of the two H3K56ac sites in (H3–H4)₂. We show that the Rtt106–(H3–H4)₂ interaction is important for gene silencing and the DNA damage response.

To understand the mode of action of Rtt106, we characterized its three-dimensional (3D) structure and probed its association with histones. Rtt106 is modular (Supplementary Fig. 1), consisting of a homodimeric N-terminal domain (Rtt106DD; residues 1–42) and a double PH domain (Rtt106PH; residues 68–301) linked via a disordered region (residues 43–67) (Fig. 1a, b and Supplementary Figs 2 and 3). The 3D structure of Rtt106DD, determined using NMR spectroscopy, shows a previously undiscovered fold with each protomer adopting a V-shaped conformation consisting of two α -helices separated by a trans-proline residue (Fig. 1a, Supplementary Fig. 3 and Supplementary Table 1). The first and second α -helices of each protomer interact with the second and first α -helices of the other protomer, respectively, through extensive hydrophobic contacts (Fig. 1a). The 3D structure of Rtt106PH, determined by X-ray crystallography to a resolution of 1.4 Å (Fig. 1b and Supplementary Table 2), reveals similarity to the structure of Pob3, a protein thought to have a role in histone deposition¹¹ (Supplementary Fig. 4).

We performed isothermal titration calorimetry (ITC) measurements to probe the interaction of Rtt106DD–Rtt106PH (residues 1–301) with the histone H3–H4 complex reconstituted using non-acetylated H3 and H3 with an acetyl-lysine analogue chemically installed at position 56 (ref. 12). Rtt106 binds both non-acetylated and K56-acetylated H3–H4. However, acetylation results in enhanced affinity (Fig. 1c). In first approximation, the biphasic curve for the interaction of Rtt106 with non-acetylated H3–H4 in Fig. 1c can be

interpreted as two concurrent binding isotherms in a two-site binding model with dissociation constants $K_{d1} = 0.4 \pm 0.2 \mu\text{M}$ and $K_{d2} = 1.5 \pm 0.2 \mu\text{M}$. Agreeing well with two binding sites, one from Rtt106DD and the other from Rtt106PH, Rtt106DD (residues 1–42) alone binds H3–H4 in an acetylation-independent manner. ITC data are consistent with a one-site binding model with a K_d of $0.6 \pm 0.1 \mu\text{M}$ (Fig. 1d), close to K_{d1} above. The reaction stoichiometry indicates that Rtt106DD, a dimer, binds two H3–H4 molecules, most likely in the form of an (H3–H4)₂ tetramer (Fig. 1d). Binding of Rtt106DD with H3–H4 was also demonstrated by NMR spectroscopy (Supplementary Fig. 5).

To interpret the biphasic ITC thermograms for the interaction of Rtt106 with K56-acetylated H3–H4, we considered a two-site binding model with an activation term that accounts for the effect of acetylation. The first dissociation constant $K_{d1ac} = 0.8 \pm 0.4 \mu\text{M}$ is similar to K_{d1} (Fig. 1c and Supplementary Fig. 6). The second dissociation constant, K_{d2ac} , is $0.08 \pm 0.06 \mu\text{M}$. In comparison to K_{d2} , the apparent gain in affinity for Rtt106PH is approximately 15–20-fold. These results indicate that the acetylated region of H3 is recognized by Rtt106PH but not by Rtt106DD. This was verified by monitoring NMR chemical shift perturbations in ¹H–¹⁵N heteronuclear single-quantum coherence (HSQC) correlation spectra of ¹⁵N-labelled Rtt106 (residues 1–67) and Rtt106PH upon titration with a non-labelled H3K56ac peptide (residues 51–61) (Fig. 1b). No binding to Rtt106 (residues 1–67) was observed upon addition of up to 15 molar excess of peptide (data not shown). In contrast, the H3K56ac peptide does specifically bind Rtt106PH as demonstrated by marked chemical shift changes ($\Delta\delta \geq 0.2$ p.p.m.) for 38 backbone amide signals of Rtt106PH (Fig. 1e and Supplementary Fig. 7). The affected residues are mapped to the second PH domain, specifically at the interface of the carboxy-terminal α -helix (α 5), two underlying β -sheets and the flexible tether connecting the two PH domains (Fig. 1b). Noticeably, this region differs from the binding sites of previously characterized PH domains¹³ (Supplementary Fig. 8). The K_d for the Rtt106PH–H3K56ac peptide interaction is 0.9 ± 0.1 mM (Supplementary Fig. 9). The markedly reduced affinity compared to that obtained for Rtt106DD–Rtt106PH (residues 1–301) and full-length H3–H4 is expected because the complete interaction is combinatorial with Rtt106DD and Rtt106PH both contacting H3–H4. Rtt106PH by itself has limited selectivity towards acetylation with only a twofold decrease in affinity for the non-acetylated H3K56 peptide ($K_d = 1.9 \pm 0.4$ mM).

How then does K56 acetylation specifically enhance the affinity of H3–H4 for Rtt106? Recent biophysical studies have highlighted the structural heterogeneity of H3 α N helix in the context of (H3–H4)₂, suggesting that the structure and dynamics of α N could be readily altered by post-translational modifications¹⁴. K56, being the C-terminal residue of α N, contributes favourably to helical stability via charge interaction with the α -helix dipole. Therefore, in neutralizing the charge of K56, acetylation may increase the conformational entropy

¹Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, Minnesota 55905, USA. ²Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, Minnesota 55905, USA.

*These authors contributed equally to this work.

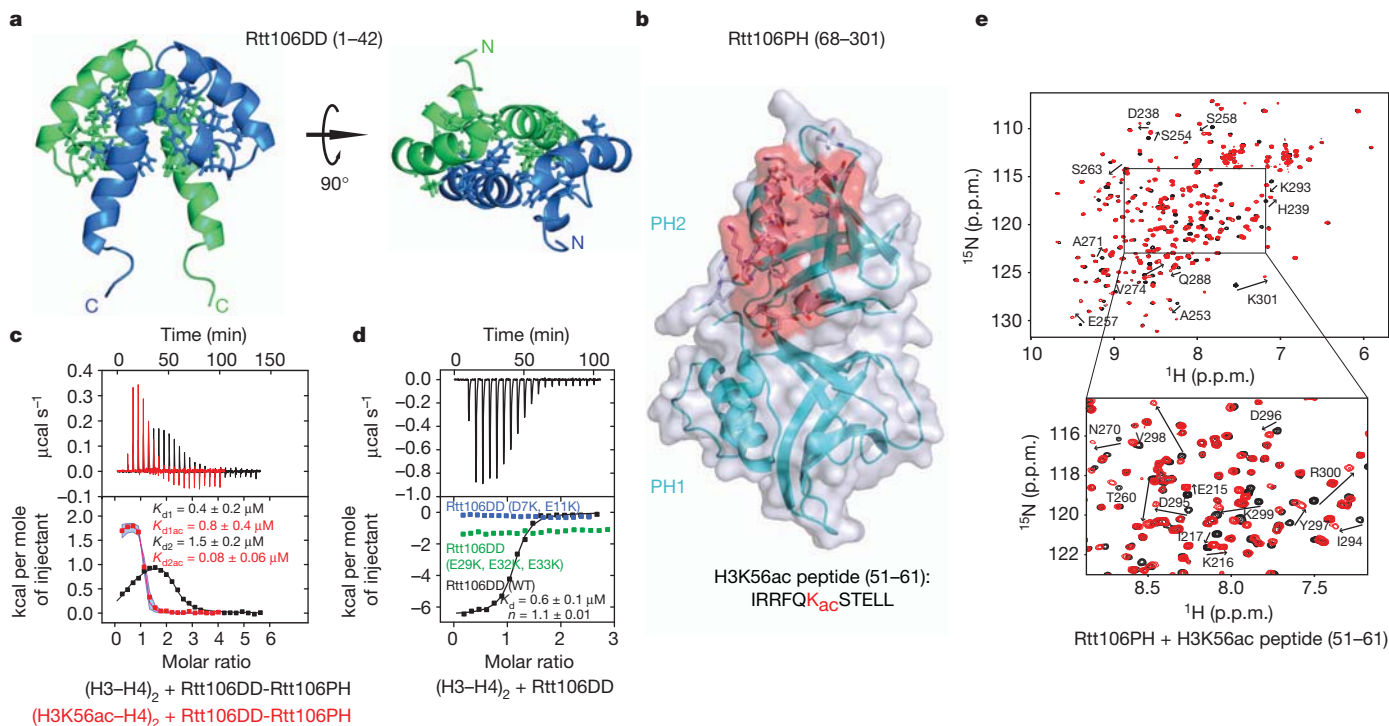


Figure 1 | 3D structures of Rtt106 dimeric and double PH domains and their interaction with histones. **a**, NMR structure of Rtt106 dimeric region (residues 1–67) with the hydrophobic residues constituting the dimerization interface in stick representation. Two protomers are in blue and green. Residues 43–67 are disordered and omitted for clarity. **b**, Crystal structure of Rtt106 double PH domain (Rtt106PH, residues 68–301) with H3K56ac peptide-binding surface in flesh. **c**, ITC results (top, raw titration data; bottom, integrated heat measurements) for the interaction of Rtt106DD–Rtt106PH (residues 1–301) with non-acetylated (H3–H4)₂ (black) and K56-acetylated (H3–H4)₂ (red). For the former interaction, a two-site binding model

(dissociation constants K_{d1} and K_{d2}) was used. For the latter, an activation step accounting for the effect of acetylation was incorporated in the two-site binding model (K_{d1ac} and K_{d2ac}). K_{dS} are reported with s.d. determined by nonlinear least-squares analysis. The light blue envelop represents simulated data for K_{d2ac} varying from 0.01 to 0.1 μM and $K_{d1ac} = 0.4 \mu\text{M}$. **d**, ITC data for the interaction of Rtt106DD (residues 1–42) with $(\text{H3-H4})_2$. Stoichiometry n and K_d are indicated. Data for two mutant forms of Rtt106DD, D7K, E11K and E29K, E32K, E33K, are also shown. **e**, ^1H - ^{15}N HSQC spectra of H3K56ac peptide-bound (red) versus free (black) Rtt106PH. Perturbed signals are labelled on the spectra.

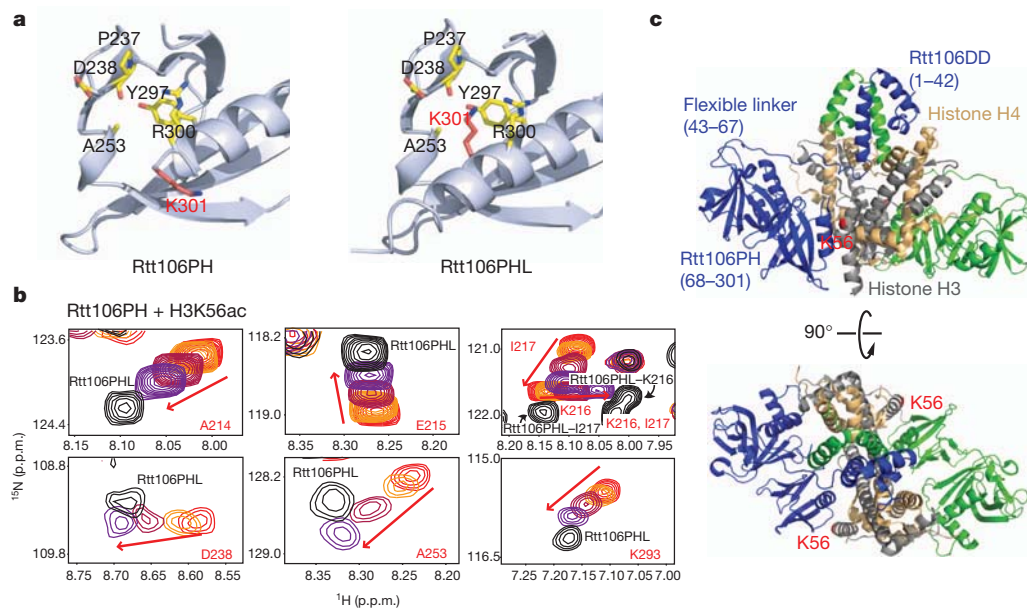


Figure 2 | Identification of a K56ac-binding cleft in Rtt106 and model of Rtt106 in complex with K56-acetylated (H3-H4)₂. **a**, Binding cleft and the side chain of K301 (red) in Rtt106PH and Rtt106PHL. **b**, Chemical shift changes in Rtt106PH ¹H-¹⁵N HSQC spectra upon titration with the H3K56ac peptide (from red to purple signals) are compared to the chemical shifts of free

Rtt106PHL (black signals) with respect to seven residues in the vicinity of the binding cleft. A214, E215, K216 and I217 belong to the disordered loop adjacent to the binding cleft. c, Structural model of Rtt106 (residues 1–301) in complex with the (H3–H4)₂ tetramer. Atomic coordinates of (H3–H4)₂ are from the structure of budding yeast nucleosome core particle (PDB access code 1ID3)⁴⁶.

of α N, favouring interaction with Rtt106. This may be correlated to the observed entropy-driven increase in affinity upon acetylation (Supplementary Fig. 6).

In comparison to Rtt106PH (residues 68–301), the crystal structure of the longer Rtt106PHL (residues 68–315), with an α -helical extension from residues 302 to 311 (Supplementary Fig. 4), reveals a change in the conformation of K301 that leads to the identification of a K56/K56ac binding cleft in Rtt106 (Fig. 2a and Supplementary Table 2). Via an approximately 180° hinged motion, the side chain of K301 changes from being solvent-exposed in the structure of Rtt106PH to being partially buried in a cleft that is part of the H3K56ac peptide-binding region identified using NMR spectroscopy (Figs 1b, e and 2a). Noticeably, the ^1H – ^{15}N chemical shifts of several Rtt106PHL residues in the vicinity of the K301-occupied cleft match well the end-point chemical shifts of corresponding Rtt106PH residues upon titration with the H3K56ac peptide (Fig. 2b and Supplementary Fig. 10). With only one lysine (K56) in the H3K56ac and H3K56 peptides (Fig. 1b), the correspondence in NMR chemical shifts between the Rtt106PH–H3K56ac complex and Rtt106PHL strongly suggests that the K301-binding cleft is also the binding pocket for acetylated and non-acetylated K56.

The structural difference between Rtt106PH and Rtt106PHL likely reflects a dynamic exchange between an open and a closed state of the binding site. Consistent with dual states, H3K56ac peptide binding to Rtt106PHL occurs, but with lower affinity than for Rtt106PH (Supplementary Fig. 11). The relatively large crystallographic B-factor values for the C-terminal residues of Rtt106PHL (residue 299 and onwards) are also consistent with conformational flexibility (Supplementary Fig. 12). Moreover, in another crystal structure of Rtt106 (residues 65–320)¹⁵, there is no detectable electron density for residues 303–320 encompassing the helical extension of Rtt106PHL. Also supporting a two-state binding site with the open conformation favouring histone binding, four mutations at the C-terminal end of Rtt106PH markedly increased Rtt106PH affinity for the H3K56ac peptide (for example, $K_d = 0.4 \pm 0.1$ mM for K299A, Supplementary Table 3).

To illustrate how Rtt106 may associate with (H3–H4)₂, we derived a structural model of the complex (Fig. 2c). In the model, in which the dyad symmetry axes of the Rtt106DD and (H3–H4)₂ structures coincide, Rtt106DD is placed in a positively charged cavity formed between the two H3–H4 subcomplexes¹⁶ and without any contact with the region of H3 encompassing K56, in accordance with experimental data. Furthermore, we verified by ITC that two different sets of mutations (D7K and E11K; and E29K, E32K and E33K) that reverse negatively charged surface areas of Rtt106DD without affecting the 3D structure disrupt binding to H3–H4 (Fig. 1d). We also note that removal of the dimeric domain renders Rtt106 non-functional *in vivo* (data not shown). In addition, in the model the flexible Rtt106 inter-domain linker (residues 43–67) has an appropriate length to position Rtt106PH in contact with the K56-containing region of H3.

The histone-binding surface of Rtt106PH was further validated by measuring the affinity of twenty Rtt106PH mutants for the H3K56ac peptide (Fig. 3a, Supplementary Table 3, Supplementary Fig. 9 and Supplementary Discussion). Several of the mutations that affect binding were then incorporated into full-length Rtt106 to assess interactions with intact histones *in vivo*. Wild-type and mutant Rtt106 proteins were produced from budding yeast, isolated by tandem affinity purification and probed for association with histones by western blot. Whereas histone H3 co-purified with wild-type Rtt106, several surface mutations introduced in Rtt106 blocked (Y261A, F269A, Y291A and I294A) or diminished (I259A and Q288A) histone binding *in vivo* (Fig. 3b). Also consistent with the *in vitro* binding data (Fig. 3a and Supplementary Table 3), reduced amounts of H3 were detected with Rtt106 harbouring the Y297A mutation in the putative K56ac binding cleft (Supplementary Fig. 13a). These results indicate that the interaction interface identified *in vitro* from titration of Rtt106PH with

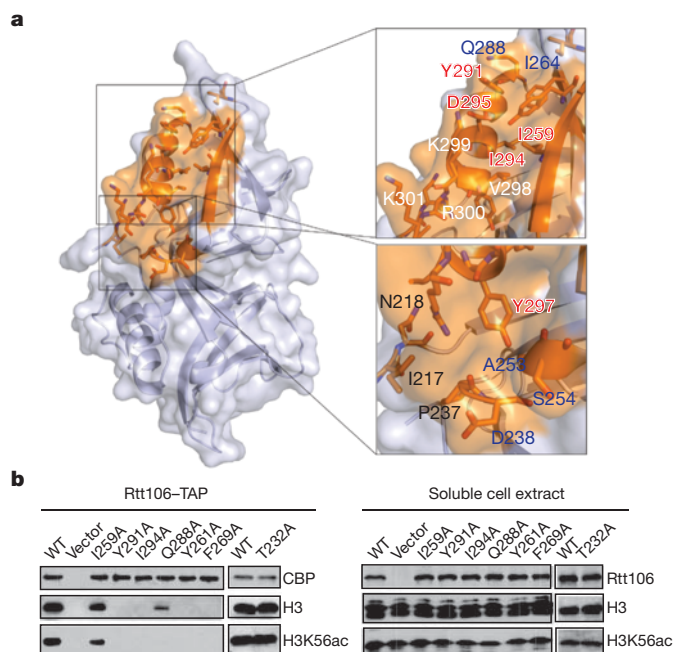


Figure 3 | Effects of Rtt106PH mutations on H3K56ac interaction.

a, Surface representation of Rtt106PH with the H3K56ac peptide-binding region in orange. The upper box represents a mainly hydrophobic area of the binding site whereas the lower box highlights the K56/K56ac binding cleft. Affinities of Rtt106PH mutants for the H3K56ac peptide were measured and K_d s are reported in Supplementary Table 3. Mutated residues that totally abolish, decrease, have no effect or enhance binding are labelled in red, blue, black and white, respectively. **b**, Wild-type (WT) and mutant tandem affinity purification (TAP)-tagged full-length Rtt106 were purified from yeast cells and analysed by western blot using indicated antibodies. CBP, calmodulin-binding peptide tag. Rtt106 mutated outside the binding site (T232A) was used as control.

an H3K56ac peptide is important for the proper interaction of Rtt106 and H3 *in vivo*.

Rtt106 is crucial for heterochromatin silencing in yeast in the absence of the Cac1 (also known as Rfl2) subunit of CAF-1, another histone chaperone implicated in K56ac-dependent replication-coupled chromatin assembly^{5,17}. Using the green fluorescent protein (GFP) as a reporter in a gene silencing assay¹⁸, we confirmed that there was significant loss of silencing of the *GFP* gene in *cac1Δrtt106Δ* cells (Fig. 4a and Supplementary Fig. 13b). *GFP* silencing was restored to almost the level in control W303-1A cells by expressing wild-type Rtt106 but not by expressing Rtt106 mutants (Y261A, F269A, Y291A and I294A) that are highly defective in H3 binding *in vivo*. Expression of Rtt106 mutants (I259A, Q288A and Y297A) that showed reduced H3K56ac binding *in vivo* slightly reduced *GFP* silencing in *cac1Δrtt106Δ* cells compared to wild-type Rtt106 expression (Fig. 4a and Supplementary Fig. 13b). These results show that the ability of Rtt106 to contact the K56-containing surface of H3 via the double PH domain is important for transcriptional silencing.

Rtt106 is also critical for maintenance of genomic stability in the absence of Cac1 (ref. 5). To test if the Rtt106 mutants that showed defects in H3 binding *in vivo* (Fig. 3b and Supplementary Fig. 13a) would have increased DNA damage sensitivity, yeast cells harbouring wild-type or mutant Rtt106 but lacking Cac1 were grown in media containing the genotoxic agents methyl methanesulphonate (MMS) or camptothecin (CPT). Rtt106 mutants severely defective for H3 interaction were more susceptible to MMS and CPT treatment than wild-type or Rtt106 mutants having little or no defect in H3 binding, indicating that the ability of Rtt106 to bind H3K56ac is connected to its role in preserving genomic integrity (Fig. 4b and Supplementary Fig. 13c).

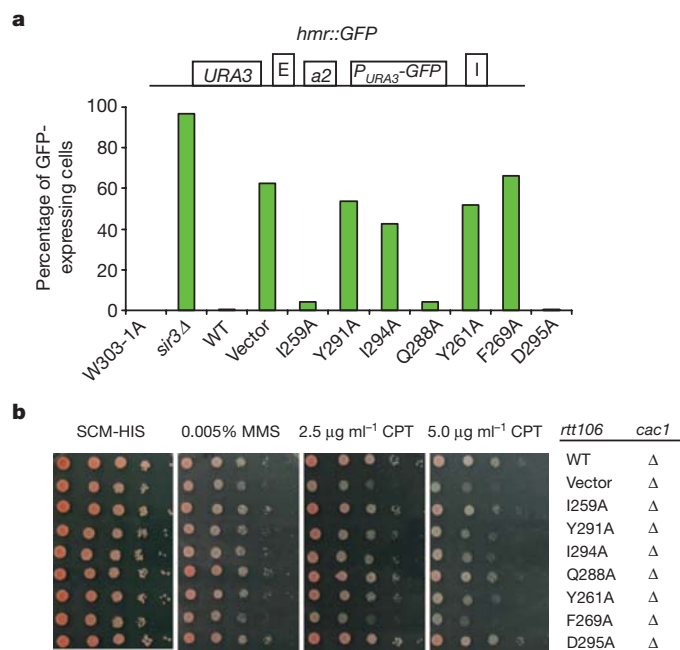


Figure 4 | Effects of Rtt106 mutations on HMR silencing and genome stability. **a**, Schematic of the GFP-based gene silencing reporter assay. The GFP gene (*P_{URA3}-GFP*) at the silent mating type locus *HMR* (*hmr::GFP*) is controlled by the *URA3* gene promoter, silencers E and I, and *a2* gene. Gene silencing is reported as percentage of yeast cells expressing GFP. One representative of three independent experiments is shown. Silencing is observed in W303-1A control cells, but not in cells lacking the silent information regulator gene *SIR3*. Expression of wild-type (WT) Rtt106 in *cac1Δrtt106Δ* cells restores silencing. Expression of Rtt106 mutated in the H3K56ac-binding surface does not or only partially restores silencing. **b**, Mutations in the H3 binding sites of Rtt106 enhance the DNA-damage sensitivity of *cac1Δ* mutant cells. Cells of the indicated genotypes were spotted onto media lacking histidine (SCM-HIS) for plasmid selection, without or with MMS or CPT for DNA damage assessment.

In conclusion, our study supports a working model where direct binding of Rtt106 to H3K56-acetylated (H3–H4)₂ tetramers contributes to nucleosome assembly with implications for DNA replication, gene silencing and maintenance of genomic stability. Our findings strongly suggest that the preferential association of Rtt106 with acetylated (H3–H4)₂ originates primarily from a K56 acetylation-triggered increase in conformational entropy of H3 αN. This mode of specific association with a modified histone is fundamentally different from that of so-called histone mark reader domains¹⁹.

METHODS SUMMARY

Wild-type and mutant Rtt106 proteins were expressed in *Escherichia coli* as His₆-fusions and purified by immobilized metal affinity and gel filtration chromatography. Purification and reconstitution of H3–H4 followed established procedures²⁰. The homogeneous site-specific installation of an acetylated lysine analogue (methylthiocarbonyl-thiaLys) in place of H3K56 in H3–H4 was done as reported¹² (Supplementary Fig. 14). Protein labelling with selenomethionine (SeMet) for X-ray crystallography studies and with ¹⁵N, ¹⁵N/¹³C and ¹⁵N/¹³C/²H for NMR spectroscopy studies was achieved by growing *E. coli* cells in SeMet- and isotope-enriched media following standard procedures. All proteins were crystallized at 15 °C. X-ray diffraction data were collected on-site and at the Advance Photon Source (APS) synchrotron facility, Argonne National Laboratory. NMR experiments were collected at 25 °C using a Bruker Avance 700 MHz spectrometer. The solution NMR structure of Rtt106 (residues 1–67) was determined using a simulated annealing-based protocol. Tandem affinity purification, gene silencing assays and DNA damage assays were done as reported^{5,17,18}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 August 2011; accepted 16 January 2012.

Published online 5 February 2012.

- Luger, K. Dynamic nucleosomes. *Chromosome Res.* **14**, 5–16 (2006).
- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
- Corpet, A. & Almouzni, G. Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol.* **19**, 29–41 (2009).
- Chen, C. C. *et al.* Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell* **134**, 231–243 (2008).
- Li, Q. *et al.* Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* **134**, 244–255 (2008).
- Williams, S. K., Truong, D. & Tyler, J. K. Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc. Natl Acad. Sci. USA* **105**, 9000–9005 (2008).
- Xie, W. *et al.* Histone H3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol. Cell* **33**, 417–427 (2009).
- Das, C., Lucia, M. S., Hansen, K. C. & Tyler, J. K. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature* **459**, 113–117 (2009).
- Tjeertes, J. V., Miller, K. M. & Jackson, S. P. Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J.* **28**, 1878–1889 (2009).
- Yuan, J., Pu, M., Zhang, Z. & Lou, Z. Histone H3–K56 acetylation is important for genomic stability in mammals. *Cell Cycle* **8**, 1747–1753 (2009).
- VanDemark, A. P. *et al.* The structure of the yFACT Pob3-M domain, its interaction with the DNA replication factor RPA, and a potential role in nucleosome deposition. *Mol. Cell* **22**, 363–374 (2006).
- Huang, R. *et al.* Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J. Am. Chem. Soc.* **132**, 9986–9987 (2010).
- Blomberg, N., Baraldi, E., Nilges, M. & Saraste, M. The PH superfold: a structural scaffold for multiple functions. *Trends Biochem. Sci.* **24**, 441–445 (1999).
- Bowman, A., Ward, R., El-Mkami, H., Owen-Hughes, T. & Norman, D. G. Probing the (H3–H4)₂ histone tetramer structure using pulsed EPR spectroscopy combined with site-directed spin labeling. *Nucleic Acids Res.* **38**, 695–707 (2010).
- Liu, Y. *et al.* Structural analysis of Rtt106p reveals a DNA binding role required for heterochromatin silencing. *J. Biol. Chem.* **285**, 4251–4262 (2010).
- White, C. L., Suto, R. K. & Luger, K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* **20**, 5207–5218 (2001).
- Huang, S. *et al.* Rtt106p is a histone chaperone involved in heterochromatin-mediated silencing. *Proc. Natl Acad. Sci. USA* **102**, 13410–13415 (2005).
- Huang, S., Zhou, H., Tarara, J. & Zhang, Z. A novel role for histone chaperones CAF-1 and Rtt106p in heterochromatin silencing. *EMBO J.* **26**, 2274–2283 (2007).
- Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
- Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304**, 3–19 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to N. Juranic, S. Macura and T. Burghardt for experimental advice, Y. Kim for assistance with X-ray data collection, Z. Zhang for suggestions on chemical synthesis, and K. Luger and A. Hieb for advice on the preparation of histones. We acknowledge the use of synchrotron beamlines 19BM and 19ID of the Structural Biology Center at Argonne National Laboratory's APS, supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-ENG-38. This work was funded in part by National Institutes of Health grants to Z.Z. and G.M.

Author Contributions Q.H. carried out the NMR spectroscopy experiments, prepared methylthiocarbonyl-aziridine and acetylated H3–H4, and performed the ITC measurements and analysis with assistance from G.M.; D.S. obtained crystals of the Rtt106 constructs, performed the X-ray diffraction measurements and solved the structures. Q.L. did the *in vivo* experiments. G.C. helped with the NMR experiments and ITC data analysis. A.F. and B.A.D. purified Rtt106 for initial crystal screening. J.R.T. worked on the refinement of crystal structures. M.V.B. contributed extensively to plasmid design, mutagenesis and protein preparations. G.M. and Z.Z. supervised the research. G.M. wrote the manuscript. All authors contributed in editing the manuscript.

Author Information The atomic coordinates and structure factors or NMR restraints of Rtt106DD, Rtt106PH, Rtt106PHL and Rtt106PH-acetyl-histamine have been deposited with the Protein Data Bank under accession codes 2LH0, 3FSS, 3TVV and 3TW1, respectively. The NMR chemical shifts of Rtt106 (residues 1–67) have been deposited in the Biological Magnetic Resonance Bank with the accession code 17832. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.M. (mer.georges@mayo.edu) or Z.Z. (zhang.zhiguo@mayo.edu).

METHODS

Protein preparation. Constructs of Rtt106 encompassing amino acids 1–315, 1–301 (Rtt106DD–Rtt106PH), 1–67, 1–42 (Rtt106DD), 68–299, 68–301 (Rtt106PH) and 68–315 (Rtt106PHL) were cloned in a modified pET vector (Novagen) encoding an N-terminal His₆-tag and a tobacco etch virus (TEV) protease cleavage site. The wild-type proteins were overexpressed in *Escherichia coli* BL21(DE3) initially grown in LB broth at 37 °C to $D_{600\text{ nm}}$ of ~0.8, then transferred to 15 °C, and after 45 min, induced by 1 mM isopropyl- β -D-thiogalactoside (IPTG) for 14 to 18 h. The cells were collected by centrifugation and resuspended in 50 mM sodium phosphate buffer, pH 7.5, 300 mM NaCl and 1 mM phenylmethanesulphonylfluoride, and lysed using an Emulsiflex C5 high-pressure homogenizer (Avestin). After centrifugation, the proteins were purified by affinity chromatography with Ni²⁺-loaded NTA resin (Qiagen) according to the manufacturer's recommended protocol. The His₆-tag was cleaved with TEV protease at 4 °C overnight and further purification was performed by size exclusion chromatography using preparative Superdex 75 or 200 columns (GE Healthcare). Cleavage of the His₆-tag leaves three residues (GHM) at the N terminus of each protein.

Point mutants of Rtt106 were made following the QuickChange site-directed mutagenesis protocol (Stratagene) and were verified by DNA sequencing. All mutants were purified as described for the wild-type protein.

The preparation of isotopically labelled Rtt106 (residues 1–67), Rtt106PH and Rtt106PHL followed similar steps as above except that instead of LB broth, M9 media containing 1 g l⁻¹ ¹⁵N NH₄Cl, 4 g l⁻¹ [¹²C₆]-D-glucose and 1 g l⁻¹ ¹⁵N Isogro (Isotec) (for ¹⁵N-labelled samples); and 1 g l⁻¹ ¹⁵NH₄Cl, 2 g l⁻¹ [¹³C₆]-D-glucose and 1 g l⁻¹ ¹⁵N/¹³C Isogro (Isotec) (for ¹⁵N/¹³C-labelled samples) were used²¹. The procedure was similar for producing ²H/¹⁵N/¹³C-labelled samples but with the *E. coli* cells grown in culture media prepared with 99.9% D₂O.

For producing selenomethionine (SeMet)-enriched Rtt106PH and Rtt106PHL, a similar protocol as above was used but with protein overexpression in the methionine-auxotroph *E. coli* strain B834(DE3) (Novagen) grown in M9 media with SeMet and all amino acids except methionine²².

The preparation of histones H3 and H4 is based on a published protocol²⁰. Histones H3 and H4 from *Xenopus laevis* were overexpressed in *E. coli* BL21 (DE3) Rosetta pLysS, purified separately under denaturing conditions and then combined to reconstitute the (H3–H4)₂ tetramer. For each histone, cells were grown at 37 °C to $D_{600\text{ nm}}$ of 0.6–0.8, induced with 1 mM IPTG, collected after 3 h and lysed using an Emulsiflex C5 high-pressure homogenizer (Avestin). After centrifugation, the pellet was washed with 1 M L-arginine monohydrochloride, 5 mM 2-mercaptoethanol, 10 mM Tris-HCl, pH 7.5 three times and with 1 M L-arginine monohydrochloride, 5 mM 2-mercaptoethanol, 1 M guanidine hydrochloride, 10 mM Tris-HCl, pH 7.5 once. Next, the pellet was dissolved in 10 mM dithiothreitol, 7 M guanidine hydrochloride, 20 mM Tris-HCl, pH 7.5 and centrifuged. The supernatant was dialysed several times in water containing 5 mM 2-mercaptoethanol for a period of 3 days and then lyophilized. The lyophilized solid was dissolved in 10 mM dithiothreitol, 6 M urea, 20 mM Tris-HCl, pH 7.5 (buffer A) and residual solids were spun down. The supernatant was passed through a Resource S cation exchange column (GE Healthcare) using buffer A as running buffer and eluted with a 0 to 1 M NaCl gradient. Resulting solutions of H3 and H4 were mixed at equimolar ratio and dialysed in 2 M NaCl, 1 mM EDTA, 5 mM 2-mercaptoethanol, 10 mM Tris-HCl, pH 7.5 (refolding buffer). Refolded H3–H4 was purified by size-exclusion chromatography using a Superdex 200 column (GE Healthcare) and refolding buffer as the running buffer.

Incorporation of an acetyl-lysine analogue in H3–H4. For incorporation of an acetyl-lysine analogue at position 56 of histone H3, we closely followed a published protocol¹². The single cysteine (C110) in histone H3 was replaced by an alanine and K56 was replaced by a cysteine. C56 of H3 (K56C, C110A) or H3 (K56C, C110A)–H4 was alkylated with methylthiocarbonyl-aziridine (MTCA) to generate the acetylated lysine analogue methylthiocarbonyl-thialys (Supplementary Fig. 14).

MTCA was prepared as reported previously¹². Briefly, to pre-cooled diethyl ether (30 ml) in a round-bottom flask (–80 °C, maintained with dry ice and acetone), aziridine (100 μ l, 1.93 mmol), triethylamine (0.27 ml, 1.93 mmol) and methyl chlorothiolformate (0.16 ml, 1.89 mmol) were added with stirring. The reaction was allowed to proceed for 3 h before diluting the reaction mixture with diethyl ether:H₂O (15 ml:10 ml). The mixture was next shaken in a separation funnel and the organic layer isolated and washed as follow: 0.01 M HCl (5 ml, twice), H₂O (5 ml, twice) and brine (10 ml, twice). The organic layer was dried over anhydrous MgSO₄ and then concentrated using a rotary evaporator. The product, MTCA (100–200 mg yield), was verified by NMR spectroscopy (Supplementary Fig. 14). Aziridine was purchased from ChemService. All other chemicals were purchased from Sigma-Aldrich. Commercial reagents were used as received without further purification.

Resource S-purified H3 mutant (K56C, C110A) was dialysed extensively in water with 5 mM 2-mercaptoethanol and then in water alone. Next, it was lyophilized

and then dissolved in 100 mM ammonium bicarbonate, pH 8.0, to a final concentration of ~2 mg ml⁻¹. MTCA was added to a final concentration of 50–200 mM. The reaction proceeded for ~200 min at room temperature. The modified product was verified by mass spectrometry, lyophilized and subsequently used for reconstitution with H4 as explained above (Supplementary Fig. 14). Alternatively, the acetylation reaction was carried out on refolded H3 (K56C, C100A)–H4. After purification with Superdex 200, H3 (K56C, C100A)–H4 was dialysed in 100 mM ammonium bicarbonate, pH 8.0, 300 mM NaCl and concentrated to 2–10 mg ml⁻¹. 100 mM of MTCA was added and the reaction was left at room temperature for ~200 min. Acetylated H3–H4 obtained from either method was then extensively dialysed in the final buffer of 20 mM Tris-HCl, pH 7.5, 100 mM NaCl for ITC experiments.

Isothermal titration calorimetry. Measurements were carried out at 10 °C using a VP-ITC titration calorimeter (MicroCal). All proteins were prepared in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. Rtt106DD, Rtt106 (residues 1–67) or Rtt106DD–Rtt106PH (residues 1–301) in the calorimeter injection syringe at concentrations ranging from 0.5 mM to 0.74 mM were delivered as a series of 5- to 8- μ l injections every 5 min to the reaction cell containing non-acetylated or K56-acetylated H3–H4 at concentrations of 20 μ M or 30 μ M. Measurements were paired with control experiments for heat of mixing and dilution. Data were analysed with Levenberg–Marquardt nonlinear regression using different models programmed in Origin 7.0 and in-house software. Data were also simulated using Mathematica (Wolfram Research).

Crystallization, data collection and structure determination. Crystals of SeMet-labelled Rtt106PH were grown at 15 °C by vapour diffusion of hanging drops by mixing 1 μ l of Rtt106PH at 30 mg ml⁻¹ in 20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM dithiothreitol, 10% glycerol with 1 μ l of reservoir solution containing 4% (v/v) Tacsimite, pH 5.0, 12% PEG 3350. The crystals were cryoprotected by transfer to reservoir solution supplemented with 20% glycerol for 10–15 min, and were quick-frozen in a cryoloop (Hampton Research) with liquid nitrogen.

Single anomalous diffraction data were collected for Rtt106PH at APS 19BM beamline, Argonne National Laboratory. Processing of diffraction images and scaling of the integrated intensities were performed using HKL3000 (ref. 23). Crystals were of the space group C2 with one molecule of Rtt106PH in the asymmetric unit and a Matthews coefficient of 2.32 Å³ Da⁻¹. The four Se atom positions were determined using SHELXD²⁴, followed by density modification with RESOLVE²⁵ and initial automatic model building using ARP/wARP²⁶. Model correction and refinement were undertaken using the programs COOT²⁷ and REFMAC5 (ref. 28). Resolution is 1.4 Å. For the Ramachandran geometry, 91.5% of all dihedral angles are located in most favoured regions and 8.5% in additionally allowed regions.

A complex of Rtt106PH and acetyl-histamine (AHN) was prepared by soaking the centred monoclinic crystals of SeMet-labelled Rtt106PH for 5 min in a 1 M solution of AHN (Sigma) prepared in the mother liquor and by flash-freezing in liquid nitrogen. Diffraction data were collected at 100 K using a Rigaku Microfocus 007 generator and Rigaku R-Axis IV⁺⁺ area detector. Data processing and scaling of the integrated intensities were performed using HKL2000 (ref. 29). The structure was solved by molecular replacement using PHASER³⁰ with the crystal structure of Rtt106PH as a search model. Model correction and refinement were undertaken using the programs COOT²⁷ and PHENIX³¹. Resolution is 1.8 Å. For the Ramachandran geometry, 92% of all dihedral angles are located in most favoured regions and 8% in additionally allowed regions.

Crystals of SeMet-labelled Rtt106PHL (residues 68–315) were grown at 15 °C by hanging drop vapour diffusion after mixing 1 μ l of Rtt106PHL at 30 mg ml⁻¹ in 20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM dithiothreitol, 10% glycerol with 1 μ l of reservoir solution containing 8% (v/v) Tacsimite, pH 5.0, 20% PEG 3350. The crystals were cryoprotected as for Rtt106PH above.

Diffraction data were collected at 100 K at the APS 19ID beamline, Argonne National Laboratory. Processing of diffraction images and scaling of the integrated intensities were performed using HKL3000 (ref. 23). Crystals were of the space group P2₁ with two molecules of Rtt106PHL in the asymmetric unit and a Matthews coefficient of 2.86 Å³ Da⁻¹.

The structure was solved by molecular replacement using PHASER³⁰ with the crystal structure of Rtt106PH as a search model. Model correction and refinement were undertaken using the programs COOT²⁷ and PHENIX³¹. Resolution is 2.6 Å. For the Ramachandran geometry, 86.8% of all dihedral angles are located in most favoured regions and 12.9% in additionally allowed regions.

NMR spectroscopy. NMR experiments were conducted at 25 °C using a Bruker Avance 700 MHz spectrometer equipped with a cryogenic probe. The Rtt106DD, Rtt106 (residues 1–67), Rtt106PH and Rtt106PHL protein samples (wild-type and mutants) were at concentrations of ~0.6 mM in 20 mM sodium phosphate buffer, pH 6.9, 30 mM NaCl and 5 mM dithiothreitol. 95% of backbone carbon and

nitrogen resonances of Rtt106PH and 98% of all resonances of Rtt106 (residues 1–67) were assigned from regular and transverse relaxation-optimized spectroscopy (TROSY)-based experiments³² using $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labelled or $^{13}\text{C}/^{15}\text{N}$ -labelled Rtt106 samples. NMR data were processed with NMRPipe/NMRDraw³³ and analysed with SPARKY (Goddard, T. D. & Kneller, D. G., <http://www.cgl.ucsf.edu/home/sparky/>). The solution structure of Rtt106 (1–67) homodimer was calculated and refined using the program CNS version 1.2, using a simulated annealing protocol for torsion angle dynamics³⁴. A total of 1,150 distance constraints derived from 3D ^{15}N -resolved nuclear Overhauser enhancement spectroscopy (NOESY), 3D ^{13}C -resolved NOESY, 2D NOESY and 2D ^{13}C filtered-edited NOESY spectra were used in the structure calculations. Also included were 48 hydrogen bond distance constraints derived from hydrogen–deuterium exchange measurements and NOESY data, and 112 dihedral angles derived from chemical shift index analysis of $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO and $^1\text{H}\alpha$ atoms. The 20 lowest energy conformers from 200 refined structures were selected to represent the NMR ensemble. For the well-folded part of the molecule (residues 6–42), 94.3% of all dihedral angles are located in most favoured regions and 5.4% in additionally allowed regions of the Ramachandran plot.

The interactions of non-acetylated histone H3 (H3K56) and H3K56ac (residues 51–61 for both) peptides with wild-type Rtt106PH were quantified by recording a series of ^1H – ^{15}N HSQC spectra of ^{15}N -labelled Rtt106PH at increasing concentrations of the peptides. The mutated Rtt106PH proteins were similarly titrated with the H3K56ac peptide. The dissociation constants (K_d s) were estimated from the NMR chemical shift perturbations of five to eight non-overlapping ^1H – ^{15}N resonances by nonlinear least-squares fitting of the following equation:

$$\frac{\Delta\delta}{\Delta\delta_{\max}} = 0.5 \left[\left(1 + \frac{K_d}{C_p} + M \right) - \sqrt{\left(\left(1 + \frac{K_d}{C_p} + M \right)^2 - 4M \right)} \right]$$

where M is the molar ratio of H3K56 or H3K56ac peptide to Rtt106PH, C_p , the concentration of Rtt106PH and $\Delta\delta$, the normalized chemical shift change calculated as:

$$\Delta\delta = \sqrt{(\delta_{\text{HN}})^2 + (\delta_{\text{N}})^2}$$

where δ_{HN} and δ_{N} denote the amide hydrogen and nitrogen atoms chemical shift differences, respectively, between the free and peptide-bound states for Rtt106PH. $\Delta\delta_{\max}$ is the normalized difference in chemical shifts of the free and peptide-saturated Rtt106PH.

Molecular illustrations. Molecular illustrations were prepared with PyMOL (<http://www.pymol.org/>) and MOLMOL³⁵.

Yeast strains, plasmids and antibodies. All budding yeast strains used in this study were derived from the parental W303 background strain (*leu2-3, ura3-1, his3-11, trp1-1, ade2-1, can1-100*) and are listed in Supplementary Table 4. Rtt106 constructs were tagged at their C termini with the tandem affinity purification (TAP) tag according to published procedures³⁶. Full-length Rtt106 containing its endogenous promoter and its C-terminal TAP tag was cloned into pRS313 vector and the resulting plasmid was used as a template to make Rtt106 mutants using the QuikChange site-directed mutagenesis kit (Stratagene). Mutant strains were constructed in the W303 background strain by standard yeast cloning methods³⁷. Antibodies used in this study were produced as described previously³⁸.

Binding of Rtt106 with H3 in yeast cells using tandem affinity purification.

To test the effect of Rtt106 mutations on H3 binding, wild-type and mutant Rtt106 proteins were purified from yeast cells using the TAP tag procedure and co-purified proteins were detected by western blotting with antibodies against calmodulin-binding peptide (CBP) and histones H3 and H3K56ac as described previously⁵.

Assay for silencing at the HMR locus using the GFP reporter. The silencing assay was performed as described previously¹⁷. Briefly, exponentially growing wild-type or mutant cells were collected, washed with PBS, resuspended in SCM-TRP media, and analysed by flow cytometry.

Assay for the sensitivity towards DNA-damaging agents. To analyse the sensitivity of *cac1Δ* yeast cells harbouring wild-type or mutant Rtt106 to different DNA-damaging agents, tenfold serial dilutions of freshly grown yeast cells were spotted onto selective media SCM-HIS containing different concentrations of methyl methanesulphonate (MMS: 0.001, 0.005 and 0.01% (v/v)) or camptothecin (CPT: 0, 1, 2.5 and 5 $\mu\text{g ml}^{-1}$). Plates were incubated at 30 °C for 3 days and then photographed.

21. Botuyan, M. V. *et al.* Structural basis of BACH1 phosphopeptide recognition by BRCA1 tandem BRCT domains. *Structure* **12**, 1137–1146 (2004).
22. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).
23. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
24. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
25. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37 (2003).
26. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
29. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
30. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60**, 432–438 (2004).
31. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
32. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29–65 (2000).
33. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
34. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
35. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
36. Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
37. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630 (1989).
38. Zhou, H., Madden, B. J., Muddiman, D. C. & Zhang, Z. Chromatin assembly factor 1 interacts with histone H3 methylated at lysine 79 in the processes of epigenetic silencing and DNA repair. *Biochemistry* **45**, 2852–2861 (2006).

Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR

Anirudh Prahallad^{1*}, Chong Sun^{1*}, Sidong Huang^{1*}, Federica Di Nicolantonio^{2,3*}, Ramon Salazar⁴, Davide Zecchin², Roderick L. Beijersbergen¹, Alberto Bardelli^{2,3} & René Bernards¹

Inhibition of the BRAF(V600E) oncoprotein by the small-molecule drug PLX4032 (vemurafenib) is highly effective in the treatment of melanoma¹. However, colon cancer patients harbouring the same BRAF(V600E) oncogenic lesion have poor prognosis and show only a very limited response to this drug^{2–4}. To investigate the cause of the limited therapeutic effect of PLX4032 in BRAF(V600E) mutant colon tumours, here we performed an RNA-interference-based genetic screen in human cells to search for kinases whose knockdown synergizes with BRAF(V600E) inhibition. We report that blockade of the epidermal growth factor receptor (EGFR) shows strong synergy with BRAF(V600E) inhibition. We find in multiple BRAF(V600E) mutant colon cancers that inhibition of EGFR by the antibody drug cetuximab or the small-molecule drugs gefitinib or erlotinib is strongly synergistic with BRAF(V600E) inhibition, both *in vitro* and *in vivo*. Mechanistically, we find that BRAF(V600E) inhibition causes a rapid feedback activation of EGFR, which supports continued proliferation in the presence of BRAF(V600E) inhibition. Melanoma cells express low levels of EGFR and are therefore not subject to this feedback activation. Consistent with this, we find that ectopic expression of EGFR in melanoma cells is sufficient to cause resistance to PLX4032. Our data suggest that BRAF(V600E) mutant colon cancers (approximately 8–10% of all colon cancers^{2,3,5}), for which there are currently no targeted treatment options available, might benefit from combination therapy consisting of BRAF and EGFR inhibitors.

Activating mutations in the BRAF oncogene (BRAF(V600E)) are seen in some 70% of primary melanomas⁶, some 10% of colorectal cancers⁷ and some 30–70% of papillary thyroid carcinoma^{8–10}. However, clinical responses to the highly selective small-molecule inhibitor of the BRAF(V600E) oncoprotein, PLX4032, differ widely, ranging from a response rate of approximately 80% in melanoma to only 5% in BRAF mutant colorectal cancer^{2–4}. To investigate the molecular mechanism responsible for the intrinsic resistance of BRAF(V600E) colorectal cancers (CRCs) to PLX4032, we first tested a panel of BRAF(V600E) mutant melanoma and CRC cell lines for their response to PLX4032. We found that the sensitivity of melanoma and CRC cells in both short-term (Fig. 1a) and long-term (Fig. 1b) proliferation assays *in vitro* mirrors the clinical experience, with melanoma cells being more sensitive to PLX4032 than CRC cells.

RNA interference (RNAi) genetic screens have been used successfully to identify genes that enhance a phenotype¹¹. We therefore set out to screen a short hairpin RNA (shRNA) library representing the full complement of 518 human kinases¹² (the 'kinome') and 17 additional kinase-related genes (Supplementary Table 1) for genes whose inhibition confers sensitivity to PLX4032 in BRAF(V600E) mutant CRC cells. WiDr cells were infected with the lentiviral kinome shRNA collection and cultured in the absence or presence of PLX4032 for 10 and 18 days, respectively. After this, the relative abundance of shRNA vectors was

determined by next generation sequencing of the barcode identifiers present in each shRNA vector (Fig. 1c; see Methods). We arbitrarily considered only shRNA vectors that had been sequenced at least 300 times and which were depleted at least fivefold by the drug treatment. Figure 1d shows that only very few of the 3,388 shRNA vectors in the library met this stringent selection criterion, among which were three independent shRNA vectors targeting the EGFR (see Supplementary Table 2 for all selected shRNAs). This suggested that suppression of EGFR synergizes with BRAF inhibition in these CRC cells. To validate this finding, we infected WiDr cells with each of these three EGFR shRNA vectors (all of which reduced EGFR levels; Fig. 1f) and cultured these cells with or without PLX4032 for 2 weeks. Figure 1e shows that inhibition of EGFR does not significantly affect proliferation of EGFR in WiDr cells, consistent with the clinical observations that KRAS or BRAF mutant CRC cells do not respond to EGFR-targeted monoclonal antibodies^{7,13,14}. In contrast, suppression of EGFR in combination with PLX4032 caused a marked inhibition of proliferation in WiDr cells (Fig. 1e). This suggested that BRAF(V600E) mutant CRC cells are responsive to treatment with a combination of BRAF inhibitor plus an EGFR inhibitor.

At present, two classes of anti-EGFR drugs are clinically available; these include the monoclonal antibodies cetuximab and panitumumab, and the small-molecule kinase inhibitors gefitinib and erlotinib. We found that three BRAF mutant CRC cell lines (WiDr, VACO432 and KM20) all lack a significant response to monotherapy with PLX4032, cetuximab or gefitinib. However, strong synergy was seen when PLX4032 was combined with either cetuximab or gefitinib (Fig. 2a and Supplementary Fig. 1A, C) or erlotinib (data not shown), consistent with the notion derived from the shRNA screen that EGFR inhibition is required to elicit a response to BRAF inhibition in CRC cells.

To address the molecular mechanism underlying the synergy between BRAF and EGFR inhibition in colon cancer, we tested lysates of drug-treated cells with phosphoprotein-specific antibodies that identify the activated state of components of the EGFR signalling pathway. To our surprise, we observed that treatment of all three BRAF mutant CRC cell lines with PLX4032 resulted in a strong increase in Tyr 1068 phosphorylation of EGFR, which reflects activation of the receptor (Fig. 2b and Supplementary Fig. 1B, D). This observation suggests that a powerful feedback activation of EGFR is elicited by BRAF inhibition. This feedback activation is ligand-dependent, as it does not take place in the absence of serum growth factors (Supplementary Fig. 2B). Co-treatment of these cells with a combination of PLX4032 and either cetuximab or gefitinib prevented this feedback activation of EGFR. PLX4032 treatment inhibited MEK and ERK activation downstream of BRAF but activated AKT, which acts downstream of EGFR in a pathway parallel to BRAF. We note that in all three cell lines treatment with BRAF and EGFR inhibitors caused a more complete inhibition of AKT, MEK and ERK signalling

¹Division of Molecular Carcinogenesis, Center for Biomedical Genetics and Cancer Genomics Centre, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

²Laboratory of Molecular Genetics, Institute for Cancer Research and Treatment, University of Torino, Medical School, Strada 142 Km 3.95, 10060 Candiolio, Torino, Italy. ³FIRC-IFOM Institute for Molecular Oncology, Via Adamello 16, 20100 Milan, Italy. ⁴IDIBELL, Institut Català d'Oncologia L'Hospitalet de Llobregat, Gran Via s/n, km 2, 7 08907 L'Hospitalet-Barcelona, Spain.

*These authors contributed equally to this work.

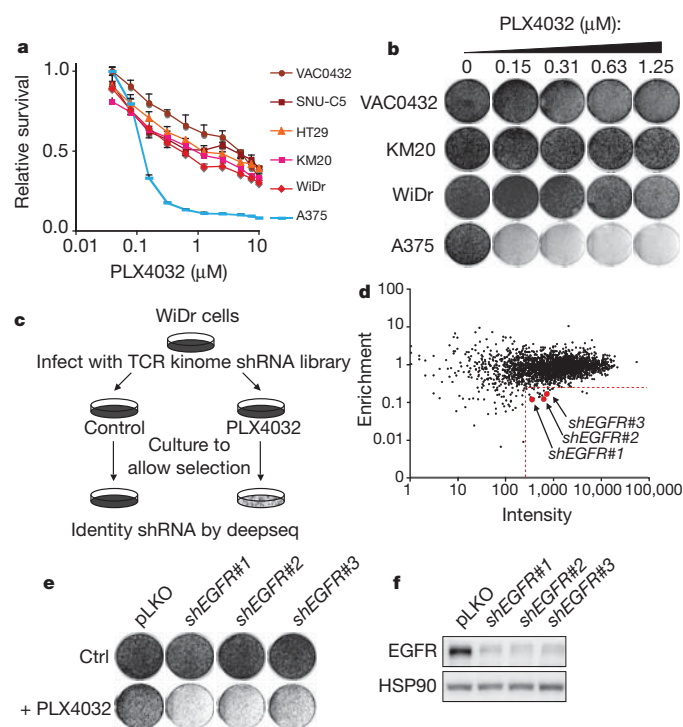


Figure 1 | EGFR inhibition confers sensitivity to BRAF(V600E) inhibition in colon cancer. **a**, **b**, CRC but not melanoma cells harbouring the BRAF(V600E) mutation are resistant to PLX4032 treatment. **a**, Short-term growth-inhibition assay of a cell line panel consisting of CRC (VACO432, SNU-C5, HT29, KM20 and WiDr) and melanoma (A375) cells. Cells were treated with increasing concentrations of PLX4032 for 72 h, and cell viability was determined using CellTiter-Blue by measuring the absorbance at 540 nm in a microplate reader. Error bars show data \pm standard error. Means were derived from four replicates ($n = 4$). **b**, Long-term colony formation assay of CRC (VACO432, KM20 and WiDr) and melanoma (A375) cells. Cells were grown in the absence or presence of PLX4032 at the indicated concentrations for 10–14 days. For each cell line, all dishes were fixed at the same time, stained and photographed. **c**, Schematic outline of the ‘dropout’ RNAi screen for enhancers of PLX4032 sensitivity. Human TRC kinome shRNA library polyclonal virus was produced to infect WiDr cells, which were then left untreated (control) for 10 days or treated with 1 μ M PLX4032 for 18 days. After selection, shRNA inserts from both populations were recovered by polymerase chain reaction (PCR) and identified by deep sequencing (deepseq). **d**, Representation of the relative abundance of the shRNA barcode sequences from the shRNA screen experiment depicted in **c**. The y axis shows enrichment (relative abundance of PLX4032 treated/untreated) and the x axis shows intensity (average sequence reads in untreated sample) of each shRNA. Among the 22 top shRNA candidates (more than fivefold depleted by PLX4032 treatment and more than 300 reads in the untreated condition (as indicated by the red dashed lines), three independent *shEGFR* vectors (in red) were identified. **e**, **f**, Three independent shRNAs targeting *EGFR* enhance response to PLX4032. **e**, The functional phenotypes of non-overlapping *shEGFR* vectors are indicated by colony formation assay in 1 μ M PLX4032. The pLKO vector was used in the control experiment (Ctrl). The cells were fixed, stained and photographed after 14 days. **f**, The level of knockdown of *EGFR* by each of the shRNAs was measured by examining the *EGFR* protein levels by western blotting.

as compared to PLX4032 monotherapy, providing a rationale for the observed synergy in growth assays (Fig. 2b and Supplementary Fig. 1B, D).

To begin to address how inhibition of BRAF(V600E) causes activation of EGFR, we treated all three BRAF mutant CRC cells with the selective MEK inhibitor AZD6244. Figure 2c shows that this drug activated EGFR (as judged by phosphorylated EGFR, p-EGFR) to the same extent as PLX4032, indicating that MEK acts downstream of BRAF to mediate the feedback regulation of EGFR. Consistent with this, we observed that expression of an active mutant of MEK (*MEK-DD*) in WiDr or VACO432 cells prevented the activation of EGFR by

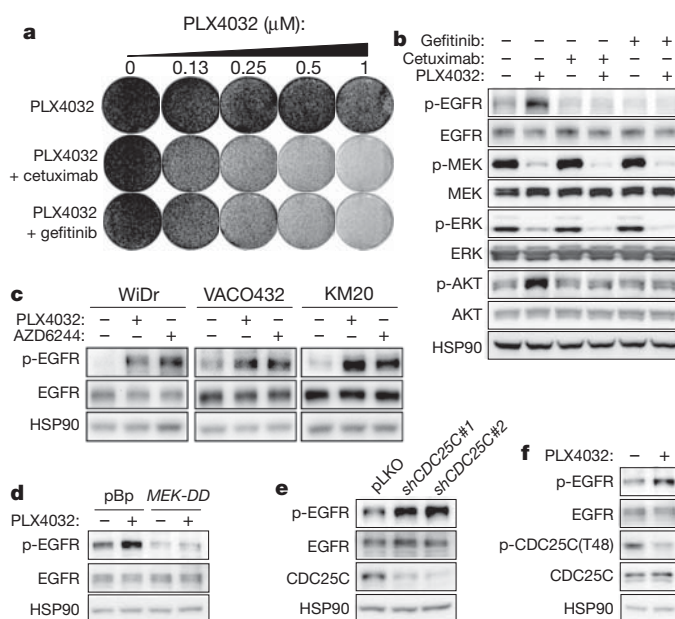


Figure 2 | Functional and biochemical interaction between BRAF and EGFR inhibition in colon cancer. **a**, Synergistic response of WiDr cells to the combination of EGFR and BRAF(V600E) inhibitors. WiDr cells were cultured in increasing concentrations of BRAF inhibitor PLX4032 alone, EGFR inhibitors cetuximab (1.25 mg ml⁻¹) or gefitinib (0.125 μ M) alone, or their combinations. The cells were fixed, stained and photographed after 18 days. **b**, Resistance to BRAF(V600E) inhibition in WiDr cells is mediated through feedback activation of EGFR. Biochemical responses of WiDr cells treated with PLX4032, cetuximab or gefitinib, or their combinations, were documented by western blot analysis. Cells were harvested at 6 h after drug treatment. BRAF(V600E) inhibition results in strong upregulation of Tyr 1068 p-EGFR and Ser 473 phosphorylated-AKT (p-AKT), which is abrogated by EGFR inhibitors. Furthermore, combination treatments result in complete inhibition of phosphorylated MEK (p-MEK) and phosphorylated ERK (p-ERK). Heat shock protein 90 (HSP90) served as a control. **c**, **d**, MEK acts downstream of BRAF to mediate the feedback regulation of EGFR in BRAF mutant CRC cells. **c**, MEK inhibitor activates p-EGFR to the same extent as PLX4032. Activation of EGFR in WiDr, VACO432 and KM20 cells treated with PLX4032 or AZD6244 for 6 h was analysed by western blot. **d**, *MEK-DD* prevents the activation of EGFR by PLX4032. Western blot analysis of EGFR in WiDr cells expressing pBabe-*PUR*O (pBp) vector control or *MEK-DD* treated with PLX4032 for 1 h. **e**, Western blot analysis showing that suppression of CDC25C by two independent shRNA vectors results in elevated levels of p-EGFR in WiDr cells. **f**, PLX4032 treatment leads to a reduced activation of CDC25C. Feedback regulation of CDC25C and EGFR in VACO432 cells treated with PLX4032 for 1 h were documented by western blot analysis.

PLX4032 (Fig. 2d and data not shown). In agreement with a central role for MEK in mediating the feedback activation of EGFR, the combination of MEK and EGFR inhibitors also synergized to inhibit growth of VACO432 or WiDr cells (Supplementary Fig. 3 and data not shown). It is unclear how MEK inhibition leads to activation of EGFR through increased phosphorylation of Tyr 1068. It has been shown in *Xenopus* that ERK kinase can phosphorylate Cdc25c on several residues, including Thr 48, leading to activation of its phosphatase activity¹⁵. Moreover, Cdc25a can bind to and dephosphorylate EGFR¹⁶. We therefore began by investigating a potential role of CDC25C in the activation of EGFR. We suppressed CDC25C in WiDr cells by shRNA and monitored levels of p-EGFR. We found that two independent *shCDC25C* vectors caused an increase in p-EGFR (Fig. 2e). Moreover, treatment of WiDr cells with PLX4032 inhibited phosphorylation of CDC25C at Thr 48 (Fig. 2f), which has been shown to be required for its phosphatase activity¹⁵. Together, these data are consistent with a model in which BRAF inhibition leads to inhibition of MEK and ERK kinases, which in turn leads to a reduced activation of CDC25C. Inhibition of CDC25C in turn causes an increase in p-EGFR

due to decreased dephosphorylation (Fig. 2e). Our data do not exclude the related CDC25A, CDC25B, or other phosphatases being involved in this feedback regulation of EGFR.

The EGFR is expressed primarily in epithelial cancers¹⁷. Because melanomas are derived from the neural crest, we reasoned that the favourable response of melanomas to PLX4032 might result from the paucity of EGFRs on these tumours and hence the absence of the feedback activation of EGFR by BRAF inhibition. We compared EGFR expression in a panel of *BRAF(V600E)* mutant melanoma, colon cancer and thyroid cancer cells. Melanoma cell lines indeed express low levels of EGFR (Fig. 3a and data not shown). Of the ten colorectal cancer cell lines examined, eight express much higher levels

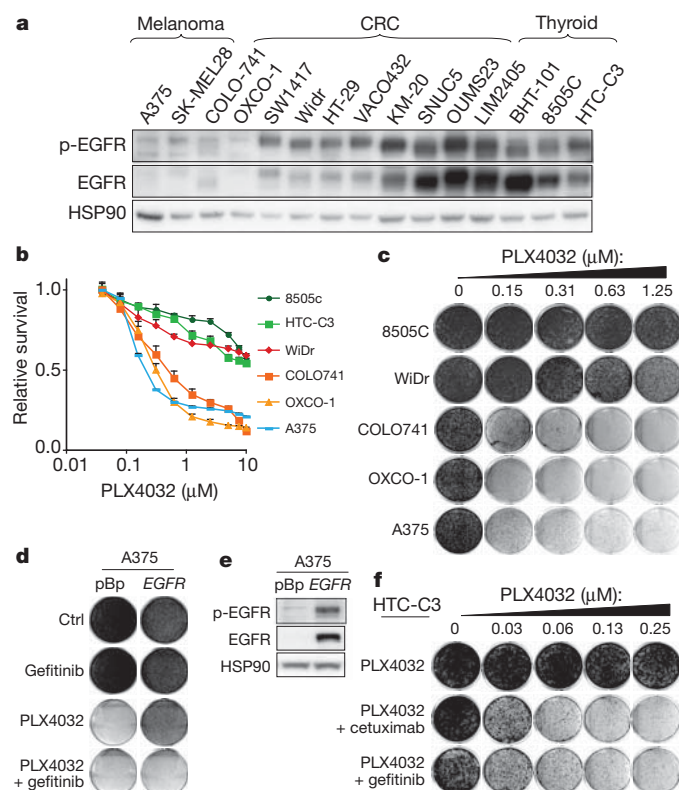


Figure 3 | Correlation between EGFR levels and response to BRAF inhibition in melanoma and CRC. **a**, Western blot analysis of p-EGFR and EGFR levels in a panel of *BRAF(V600E)* mutant cell lines from melanoma, CRC and thyroid cancer. HSP90 served as a control. **b**, High levels of EGFR expression in tumour types harbouring *BRAF(V600E)* mutations correlate with PLX4032 resistance. Short-term growth-inhibition assays of a cell line panel consisting of thyroid cancer (HTC-C3 and 8505C), CRC (OXCO-1, COLO741 and WiDr) and melanoma (A375) cells. Cells were treated with increasing concentrations of PLX4032 for 72 h, and cell viability was determined using CellTiter-Blue by measuring the absorbance at 540 nm in a microplate reader. Error bars show data \pm standard error. Means were derived from four replicates ($n = 4$). **c**, Long-term colony formation assay of thyroid cancer (8505C), CRC (OXCO-1, COLO741 and WiDr) and melanoma (A375) cells. Cells were grown in the absence or presence of PLX4032 at the indicated concentrations for 10–14 days. For each cell line, all dishes were fixed at the same time, stained and photographed. **d**, Ectopic EGFR expression confers resistance to PLX4032, but not to the combination of PLX4032 and gefitinib in A375 melanoma cells. A375 cells expressing pBabe-PURO (pBp) vector control or *EGFR* were cultured in PLX4032 (5 μ M), gefitinib (2.5 μ M) or their combination. The cells were fixed, stained and photographed after 7 (untreated or gefitinib) or 9 (PLX4032 alone or in combination with gefitinib) days. **e**, Western blot analysis of p-EGFR and total EGFR levels in cells described above. HSP90 served as a control. **f**, Synergistic response of thyroid cancer HTC-C3 cells to the combination of EGFR and BRAF(V600E) inhibitors. HTC-C3 cells were cultured in increasing concentrations of PLX4032 alone, cetuximab (1.25 mg ml⁻¹) or gefitinib (2.5 μ M) alone, or their combinations. The cells were fixed, stained and photographed after 18 days.

of EGFR, but in two (COLO-741 and OXCO-1) EGFR levels were as low as those seen in melanoma. All three thyroid cancer cell lines expressed EGFR at significant levels (Fig. 3a). When we tested the two EGFR^{low} CRC cell lines for their response to PLX4032, we found them to be almost as sensitive as the melanoma cell line A375 in both short-term assays (Fig. 3b) and long-term assays (Fig. 3c), consistent with the notion that EGFR levels determine the response to small-molecule BRAF inhibition. Thyroid cancer cell lines, which express high levels of EGFR, also responded well to the combination of gefitinib and PLX4032, but not to PLX4032 monotherapy (Fig. 3b, c, f and Supplementary Fig. 5). Finally, we tested whether ectopic expression of EGFR was sufficient to confer resistance to PLX4032 in melanoma cells. We transduced EGFR-negative A375 and SK-MEL-28 melanoma cells with a retroviral EGFR expression vector and subjected these cells to treatment with PLX4032 monotherapy or combination with the EGFR inhibitor gefitinib. Figure 3d, e and Supplementary Fig. 4 show that expression of EGFR conferred resistance to PLX4032, but these cells still responded well to the combination of PLX4032 with gefitinib. Together, these data indicate that EGFR levels in *BRAF* mutant CRC, melanoma and thyroid cancer cells are a key determinant of response to BRAF inhibitor monotherapy.

Synergistic anticancer effects of drugs can result from an effect on cell proliferation, cell death, or both. Induction of cell death is preferred to prevent the re-growth of the tumour after the therapy has been terminated. We measured induction of programmed cell death (apoptosis) by measuring the induction of cleaved poly(ADP-ribose) polymerase (PARP), a hallmark of apoptosis¹⁸. Neither PLX4032 nor EGFR inhibitors alone induced PARP cleavage in WiDr or VACO cells. However, the two drugs combined did induce marked PARP cleavage, indicative of apoptosis induction (Fig. 4a, b). This suggested that BRAF and EGFR inhibition should also be synergistic when combined *in vivo*. To test this, we used immunodeficient mice xenografted with human WiDr and VACO432 CRC tumours. Ten days after injection of tumour cells, palpable tumours were present in all animals, and cohorts of mice were treated with vehicle, the EGFR-targeted drugs cetuximab or erlotinib, the BRAF inhibitor PLX4720

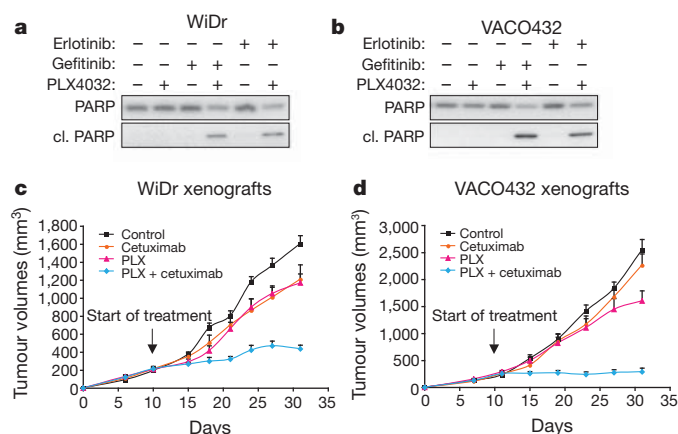


Figure 4 | EGFR and BRAF(V600E) inhibitors synergize to induce apoptosis of CRC cells and to suppress CRC tumour growth in a xenograft model. **a**, **b**, Combination of EGFR and BRAF(V600E) inhibitors leads to apoptosis in CRC cells. Western blot analysis of full-length PARP or cleaved PARP (cl. PARP) in WiDr (**a**) and VACO432 (**b**) cells treated with PLX4032, gefitinib or erlotinib, or their combinations. Cells were harvested at 48 h after drug treatment. **c**, **d**, EGFR inhibitor cetuximab in combination with BRAF(V600E) inhibitor PLX4720 significantly suppresses tumour growth in two different xenograft models. WiDr (**c**) and VACO432 (**d**) cells were grown as tumour xenografts in NOD-SCID mice. After tumour establishment (200–250 mm³), mice were treated with vehicle, cetuximab (40 mg kg⁻¹), PLX4720 (50 mg kg⁻¹) or cetuximab (40 mg kg⁻¹) plus PLX4720 (50 mg kg⁻¹), for 14 days. Mean tumour volumes \pm standard error of the mean are shown ($n = 6$ mice per group).

(highly related to PLX4032, but easier to formulate for *in vivo* use), or the combination of EGFR inhibitor drug plus PLX4720 (see Methods). Figure 4c, d shows that treatment of mice with either PLX4720 or cetuximab monotherapy resulted in marginal growth inhibition. In contrast, the combination of cetuximab and PLX4720 elicited a potent growth inhibition of WiDr and VACO432 CRC tumours. Similarly, WiDr xenografted mice treated with the combination of erlotinib and PLX4720 derived significantly more benefit than those treated with either drug alone, mirroring the observations made *in vitro* (Supplementary Fig. 6).

Our data provide a strong rationale for a clinical trial combining BRAF and EGFR inhibitors in *BRAF*(V600E) mutant CRCs, which have a very poor clinical outcome and for which no targeted therapeutic strategies are effective after failure of standard chemotherapeutic regimens^{2,3}. At first glance, it would seem counterintuitive to consider treating a *BRAF* mutant colon cancer with an EGFR inhibitor, as multiple clinical studies in colon cancer have shown that EGFR inhibition is without clinical benefit when either *KRAS* or *BRAF* is mutated downstream of EGFR^{7,13,14}. The strong synergistic interaction between inhibition of BRAF and EGFR described here is explained by an unexpected and powerful feedback activation of EGFR caused by BRAF inhibition, providing a rationale for the poor clinical response of *BRAF*(V600E) mutant colon cancers to PLX4032 monotherapy⁴. The feedback activation of EGFR also implies that *EGFR* expression levels may be a clinically useful biomarker to predict the response to PLX4032 monotherapy in *BRAF* mutant tumours. As *BRAF*(V600E) mutations are also common in thyroid papillary carcinomas and hairy-cell leukaemias, *EGFR* expression levels may also help guide the selection of EGFR combination therapy in these cancers^{8–10,19}. As some studies have shown EGFR expression in a subset of melanomas²⁰, (acquired) EGFR expression may also explain some of the clinical resistance against PLX4032 in melanoma.

There are a plethora of targeted drug agents in late-stage clinical development for the treatment of cancer. Given that resistance to monotherapy with these targeted therapies often develops rapidly, there is a trend towards combining targeted agents in clinical trials. However, the number of possible combinations of these agents seems endless. Our results highlight the power of 'synthetic lethality' genetic screens to identify which combinations of pathway inhibition are particularly effective. As such, these screens may help prioritize which combination therapies have the highest likelihood of being successful in the clinic.

METHODS SUMMARY

Pooled 'dropout' shRNA screen. A kinome shRNA library targeting the full complement of 518 human kinases and 17 kinase-related genes was constructed from the The RNAi Consortium (TRC) human genome-wide shRNA collection (TRC-Hs1.0). The kinome library was used to generate pools of lentiviral shRNA to infect WiDr cells. Cells stably expressing shRNA were cultured in the presence or absence of PLX4032. Massive parallel sequencing was applied to determine the abundance of shRNA in cells. shRNAs prioritized for further analysis were selected by the fold depletion caused by PLX4032 treatment.

Mouse xenografts and *in vivo* drug studies. All animal procedures were approved by the Ethical Commission of the Institute for Cancer Research and Treatment and by the Italian Ministry of Health. WiDr cells were injected subcutaneously into the right posterior flanks of 7-week-old immunodeficient NOD-SCID female mice (6 mice per group; Charles River). Tumour formation was monitored twice a week, and tumour volume based on caliper measurements was calculated by the modified ellipsoidal formula: tumour volume = $\frac{1}{2}$ length \times width (ref. 2). When tumours reached a volume of approximately 200–250 mm³, mice were randomly assigned to treatment with vehicle or drug(s).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 September 2011; accepted 18 January 2012.

Published online 26 January 2012.

- Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with *BRAF* V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
- Roth, A. D. *et al.* Prognostic role of *KRAS* and *BRAF* in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J. Clin. Oncol.* **28**, 466–474 (2010).
- Richman, S. D. *et al.* *KRAS* and *BRAF* mutations in advanced colorectal cancer are associated with poor prognosis but do not preclude benefit from oxaliplatin or irinotecan: results from the MRC FOCUS trial. *J. Clin. Oncol.* **27**, 5931–5937 (2009).
- Kopetz, S. *et al.* PLX4032 in metastatic colon cancer patients with mutant *BRAF* tumors. *J. Clin. Oncol.* **28**, abstract 3534 (2010).
- Tol, J., Nagtegaal, I. D. & Punt, C. J. *BRAF* mutation in metastatic colorectal cancer. *N. Engl. J. Med.* **361**, 98–99 (2009).
- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Di Nicolantonio, F. *et al.* Wild-type *BRAF* is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 5705–5712 (2008).
- Nikiforova, M. N. *et al.* *BRAF* mutations in thyroid tumors are restricted to papillary carcinomas and anaplastic or poorly differentiated carcinomas arising from papillary carcinomas. *J. Clin. Endocrinol. Metab.* **88**, 5399–5404 (2003).
- Kimura, E. T. *et al.* High prevalence of *BRAF* mutations in thyroid cancer: genetic evidence for constitutive activation of the RET/PTC-RAS-BRAF signaling pathway in papillary thyroid carcinoma. *Cancer Res.* **63**, 1454–1457 (2003).
- Trovisco, V. *et al.* *BRAF* mutations are associated with some histological types of papillary thyroid carcinoma. *J. Pathol.* **202**, 247–251 (2004).
- Kaelin, W. G. Jr. The concept of synthetic lethality in the context of anticancer therapy. *Natl. Rev.* **5**, 689–698 (2005).
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
- Karapetis, C. S. *et al.* *K-ras* mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med.* **359**, 1757–1765 (2008).
- Amado, R. G. *et al.* Wild-type *KRAS* is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 1626–1634 (2008).
- Wang, R. *et al.* Regulation of Cdc25C by ERK-MAP kinases during the G₂/M transition. *Cell* **128**, 1119–1132 (2007).
- Wang, Z., Wang, M., Lazo, J. S. & Carr, B. I. Identification of epidermal growth factor receptor as a target of Cdc25A protein phosphatase. *J. Biol. Chem.* **277**, 19470–19475 (2002).
- Salomon, D. S., Brandt, R., Ciardiello, F. & Normanno, N. Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit. Rev. Oncol. Hematol.* **19**, 183–232 (1995).
- Simbulan-Rosenthal, C. M., Rosenthal, D. S., Iyer, S., Boulares, A. H. & Smulson, M. E. Transient poly(ADP-ribosylation) of nuclear proteins and role of poly(ADP-ribose) polymerase in the early stages of apoptosis. *J. Biol. Chem.* **273**, 13703–13712 (1998).
- Tiacci, E. *et al.* *BRAF* mutations in hairy-cell leukemia. *N. Engl. J. Med.* **364**, 2305–2315 (2011).
- Boone, B. *et al.* EGFR in melanoma: clinical significance and potential therapeutic target. *J. Cutan. Pathol.* **38**, 492–502 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Kumar, V. Gambino and W. Grenrum for assistance with experiments. We are grateful to C. Pochet for support. Major financial support for this work was provided by the EU Seventh Framework Programme, grant agreement 259015 (to A.B. and R.B.). Additional funding was provided by a European Research Council grant to R.B.; The Cancer Systems Biology Center grant by the Netherlands Organisation for Scientific Research (NWO); The Dutch Cancer Society; The Netherlands Genomics Initiative (NGI); The Associazione Italiana per la Ricerca sul Cancro (AIRC), 2010 Special Program Molecular Clinical Oncology 5x1000, project 9970; AIRC Investigator Grant (to A.B.); Italian Ministry of University and Research; Italian Ministry of Health, Regione Piemonte (to A.B. and F.D.N.); Intramural Grant, 5xmille 2008; and Fondazione Piemontese per la Ricerca sul Cancro (to A.B. and F.D.N.).

Author Contributions R.B. and A.B. conceived the project and supervised all research. R.B., A.P., C.S. and S.H. wrote the manuscript. S.H., R.S. and R.L.B. designed the experiments. A.P., C.S., S.H., F.D.N. and D.Z. performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.B. (r.bernards@nki.nl).

METHODS

Cell lines and reagents. A375, SW1417 and HEK293T cells were from the laboratory collection of R.B. SK-MEL-28 cells were a gift from D. Peepers. BHT-101 and HTC-C3 cells were purchased from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Germany. These cells were cultured in DMEM supplemented with 8% FCS and 1% penicillin/streptomycin. WiDr and HT-29 cells were purchased from American Type Culture Collection (ATCC). VACO432, KM-20, SNUC5, OUMS23 and LIM2405 cells were from the laboratory collection of A. Bardelli. HTC-C3, 8508C and BHT-101 cells were purchased from DSMZ. These cells were cultured in RPMI supplemented with 8% FCS and 1% penicillin/streptomycin.

PLX4032 (catalogue no. S1267), PLX4720 (catalogue no. S1152) and gefitinib (catalogue no. S1025) were purchased from Selleck Chemicals. Erlotinib (catalogue no. RP01332e) was purchased from Sequoia Chemicals. Cetuximab was obtained from the Hospital Pharmacies at The Netherlands Cancer Institutes and from the Institute for Cancer Research and Treatment.

The TRC human genome-wide shRNA collection (TRC-Hs1.0) was purchased from Open Biosystems. Further information is available at <http://www.broadinstitute.org/rnai/public/>.

shRNA 'dropout' screen with a custom TRC kinome library. Lentiviral plasmids (pLKO.1) encoding shRNAs that target kinome candidates are listed in Supplementary Table 1. The kinome library consists of seven plasmids pools (TK1-TK7). Lentiviral supernatants were generated as described at <http://www.broadinstitute.org/rnai/public/resources/protocols>. WiDr cells were infected separately by the seven virus pools (multiplicity of infection <1) and selected with puromycin ($2 \mu\text{g ml}^{-1}$) for cells expressing integrated shRNA. Cells were then pooled and plated at 300,000 cells per 15 cm dish in the absence or presence of $1 \mu\text{M}$ PLX4032 (five dishes for each condition) and the medium was refreshed twice per week for 18 days. Genomic DNA was isolated as described²¹. shRNA inserts were retrieved from 8 μg genomic DNA by PCR amplification (PCR1 and PCR2, see below for primer information) using the following conditions: (1) 98°C , 30 s; (2) 98°C , 10 s; (3) 60°C , 20 s; (4) 72°C , 1 min; (5) to step (2), 15 cycles; (6) 72°C , 5 min; (7) 4°C . Indexes and adaptors for deep sequencing (Illumina) were incorporated into PCR primers. 2.5 μl of PCR1 products were used as templates for PCR2 reaction. PCR products were purified using Qiagen PCR purification Kit according to the manufacturer's manual. Sample quantification was performed by BioAnalyzer to ensure samples generated at different conditions were pooled at the same molar ratio before analysis by the Illumina genome analyser.

The shRNA stem sequence was segregated from each sequencing read and aligned to the TRC library. The matched reads were counted and the counts were transformed to abundance that was assigned to the corresponding shRNA.

The primers used are as follows: PCR1_Untreated replicate#1, forward, ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGATCCTTGTGGAAA GGACGAAACACCGG; PCR1_Untreated replicate#2, forward, ACACTCTTT CCCTACACGACGCTCTTCCGATCTAAGCTACTTGTGGAAAGGACGAAA CACCGG; PCR1_PLX treated replicate#1, forward, ACACTCTTTCCCTACA CGACGCTCTTCCGATCTGTAGCCCTTGTGGAAAGGACGAAAACACCGG; PCR1_PLX treated replicate#1, forward, ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACAAGCTTGTGGAAAGGACGAAAACACCGG; PCR1, reverse (P7_pLKO1_r), CAAGCAGAAGACGGCATACGAGATTTCTTCCC CTGCACTGTACCC; PCR2, forward, AATGATACGCGACCAACCGAG ATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT; PCR2, reverse (P5_IlluSeq), CAAGCAGAAGACGGCATACGAGAT.

Individual shRNA vectors used were collected from the TRC library. EGFR: TRCN0000039633_GCTGAGAATGTGGAATACCTA; TRCN0000121068_GCCACAAAGCAGTGAATTTAT; TRCN0000121206_GCCAAAGCCAAATGGCATCTTT. CDC25C: TRCN0000002433_GAAGAGAATAATCATCGTGTT; TRCN0000002434_GCCTTGAGTTGCATAGAGATT.

Short-term growth inhibition assays. Cultured cells were seeded into 96-well plates (2,000–3,000 cells per well). Twenty-four hours after seeding, serial dilutions of PLX4032 were added to cells to final drug concentrations ranging from 0.04– $10 \mu\text{M}$. Cells were then incubated for 72 h and cell viability was measured using the CellTiter-Blue viability assay (Roche). Relative survival in the presence of PLX4032 was normalized to the untreated controls after background subtraction.

Long-term cell proliferation assays. Cells were seeded into 6-well plates (2×10^4 cells per well) and cultured both in the absence and presence of drugs as indicated. For the EGFR overexpression experiments, A375 and SK-MEL-28 cells were seeded $5\text{--}10 \times 10^4$ cells per well. More details are as described²². All relevant assays were performed independently at least three times.

Protein lysate preparation and immunoblots. The biochemical responses of cells treated with drugs were analysed by western blot. Cells were plated in medium containing 10% FCS. After 24 h, cells were washed with serum-free medium and cultured for 24 h in medium containing 0.1% serum. After the low serum incubation, cells were treated with drugs for 30 min and stimulated by 10% FCS. The lysates were then collected after 6 h using sample buffer containing 5% β -mercaptoethanol, 150 mM NaCl, 50 mM Tris pH 7.5, 2 mM EDTA pH 8, 25 mM NaF and 1% NP-40, protease inhibitors (Complete, Roche), Phosphatase Inhibitor Cocktails II and III (Sigma). All lysates were freshly prepared and resolved by SDS gel electrophoresis and followed by western blotting. Primary antibodies used are as follows: p-EGFR (Y1068) (AbCam), p-ERK1/2 (T202/Y204), ERK1, ERK2, HSP-90 (SantaCruz), p-MEK1/2 (S217/221), MEK1/2, pAKT(S473), ATK1/2, PARP, CDC25C (5H9) and p-CDC25C Thr 48 (Cell Signalling Technology).

Mouse xenografts and *in vivo* drug studies. All drugs for *in vivo* studies were dissolved in DMSO, stored in aliquots at -80°C and diluted daily in aqueous vehicle before administration. All animals were manipulated according to protocols approved by the Ethical Commission of the Institute for Cancer Research and Treatment and by the Italian Ministry of Health. All experiments were performed in accordance with relevant local and national guidelines and regulations. WiDr cells (6×10^6 cells per mouse) were injected subcutaneously into the right posterior flanks of 7-week-old immunodeficient NOD-SCID female mice (6 mice per group; Charles River). Tumour formation was monitored twice a week, and tumour volume based on caliper measurements was calculated by the modified ellipsoidal formula: tumour volume = $\frac{1}{2}$ length \times width (ref. 2). When tumours reached a volume of approximately 200–250 mm^3 , mice were randomly assigned to treatment with vehicle (0.2% Tween 80 and 1% methylcellulose in sterile PBS by daily gavage), cetuximab (40 mg kg^{-1} of body weight intraperitoneally twice a week), erlotinib (80 $\text{mg kg}^{-1} \text{ day}^{-1}$ in 0.2% Tween 80 and 1% methylcellulose in sterile PBS by orogastric gavage), PLX4720 (50 $\text{mg kg}^{-1} \text{ day}^{-1}$ in 0.2% Tween 80 and 1% methylcellulose in sterile PBS by orogastric gavage); or to a drug combination (cetuximab plus PLX720, or erlotinib plus PLX4720), in which each compound was administered at the same dose and scheduled as single agents.

21. Brummelkamp, T. R. *et al.* An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. *Nature Chem. Biol.* **2**, 202–206 (2006).

22. Huang, S. *et al.* ZNF423 is critically required for retinoic acid-induced differentiation and is a marker of neuroblastoma outcome. *Cancer Cell* **15**, 328–340 (2009).

Recognition of SUMO-modified PCNA requires tandem receptor motifs in Srs2

Anthony A. Armstrong¹, Firaz Mohideen¹ & Christopher D. Lima¹

Ubiquitin (Ub) and ubiquitin-like (Ubl) modifiers such as SUMO (also known as Smt3 in *Saccharomyces cerevisiae*) mediate signal transduction through post-translational modification of substrate proteins in pathways that control differentiation, apoptosis and the cell cycle, and responses to stress such as the DNA damage response. In yeast, the proliferating cell nuclear antigen PCNA (also known as Pol30) is modified by ubiquitin in response to DNA damage and by SUMO during S phase. Whereas Ub-PCNA can signal for recruitment of translesion DNA polymerases, SUMO-PCNA signals for recruitment of the anti-recombinogenic DNA helicase Srs2. It remains unclear how receptors such as Srs2 specifically recognize substrates after conjugation to Ub and Ubls. Here we show, through structural, biochemical and functional studies, that the Srs2 carboxy-terminal domain harbours tandem receptor motifs that interact independently with PCNA and SUMO and that both motifs are required to recognize SUMO-PCNA specifically. The mechanism presented is pertinent to understanding how other receptors specifically recognize Ub- and Ubl-modified substrates to facilitate signal transduction.

Post-translational modification of substrate proteins by ubiquitin (Ub) and ubiquitin-like (Ubl) modifiers such as SUMO mediates signal transduction in pathways, including those that control differentiation, apoptosis, the cell cycle and stress responses such as the DNA damage response^{1,2}. For signal transduction to occur, downstream factors must specifically recognize the Ub- and Ubl-modified substrate. Whereas interactions between Ub and Ubl binding domains and Ub and Ubls have been addressed³, there is a paucity of data addressing how receptors specifically recognize Ub- and Ubl-conjugated substrates.

Proliferating cell nuclear antigen (PCNA) is a member of the β -clamp family of DNA sliding clamps and facilitates DNA replication as a processivity factor and mobile scaffold to recruit proteins that regulate DNA metabolism⁴. The PCNA homotrimeric ring is formed by protomers that contain two topologically related domains (domains I and II) connected by the interdomain connector loop (IDCL)⁵. Many proteins interact with PCNA at the interface between the IDCL and domains I and II with a motif coined the PCNA-interacting protein box (PIP box)^{6–8}. Post-translational modification of PCNA-interacting proteins, as illustrated by p21 phosphorylation⁹, or of PCNA itself has been proposed as a means to dynamically modulate interactions with PCNA⁴.

Replication fork stalling triggers mono- or poly-ubiquitination of PCNA on an evolutionarily conserved lysine (K164 in budding yeast), leading to activation of error-prone or error-free post-replication repair (PRR) pathways, respectively^{10,11}. MonoUb-PCNA signals for recruitment of translesion DNA polymerases such as Pol η and Pol ζ ^{10,11}. PCNA K164 is also modified by SUMO in budding yeast during S phase as is K127, albeit to a lesser extent¹⁰. SUMO-PCNA signals for recruitment of Srs2 (refs 12, 13), a UvrD-like helicase that exerts antirecombinogenic functions by disrupting Rad51 presynaptic filaments^{14,15}. Srs2 recruitment to SUMO-PCNA inhibits a Rad52-dependent recombinational repair pathway and deletion of Srs2 suppresses ultraviolet- and chemically induced damage sensitivity of PRR deficient strains^{12,13,16}. K127 and K164 SUMO modifications are redundant because either can function to recruit Srs2 (refs 12, 13).

Srs2 harbours a C-terminal SUMO interaction motif (SIM) that was shown to enhance interaction with SUMO-PCNA¹², but it is unclear how Srs2 specifically recognizes SUMO-PCNA as it was not known to contain a canonical PCNA interaction motif. At least two models can be envisioned. In a dual-interaction model distinct elements within Srs2 independently recognize PCNA and SUMO. This model is distinct, albeit not mutually exclusive, from one where the Srs2 SIM is recognized by a composite surface between SUMO and PCNA. This mode of interaction is illustrated by SUMO-modified thymine DNA glycosylase (TDG) wherein the SIM is buried in the interface between TDG and SUMO¹⁷. Srs2 recruitment to SUMO-PCNA has been touted as an ideal system for analysis because the receptor (Srs2), the SUMO-modified substrate (PCNA) and the consequence of the interaction are known^{12,18}. Here we define determinants required for specific recognition of SUMO-PCNA by Srs2 to gain insight to how other Ubl-modified substrates may be recognized by their cognate receptors.

Srs2 interacts with SUMO and PCNA

A 139-residue C-terminal fragment of Srs2 (Srs2_{1036–1174}) was shown to interact with SUMO and PCNA by yeast two-hybrid analysis¹². We found a smaller domain encompassing residues 1107–1174 that comigrated with PCNA or SUMO using analytical gel filtration (Supplementary Fig. 1). Interactions were quantified using a fluorescence polarization assay with Srs2_{1107–1174} amino-terminally coupled to the dye BODIPY-FL. Srs2_{1107–1174} interacted with glutathione-S-transferase-conjugated SUMO (GST-SUMO) with an apparent dissociation constant (K_d) of 877 ± 59 nM and with PCNA with a K_d of 169 ± 26 nM (Fig. 1; Supplementary Table 1 and Supplementary Figs 2 and 3), values similar to those observed for other SIM-SUMO^{19,20} and PIP-PCNA interactions^{7,21}. These data indicate that Srs2_{1107–1174} harbours motifs that interact independently with SUMO and PCNA.

SUMO modification of PCNA was shown to enhance interaction with Srs2 based on enrichment of SUMO-PCNA over PCNA in pull-downs^{12,13}. To address this with a purified system we reconstituted SUMO_{K164}-PCNA and SUMO_{K127}-PCNA using the Siz1

¹Structural Biology Program, Sloan-Kettering Institute, New York, New York 10065, USA.

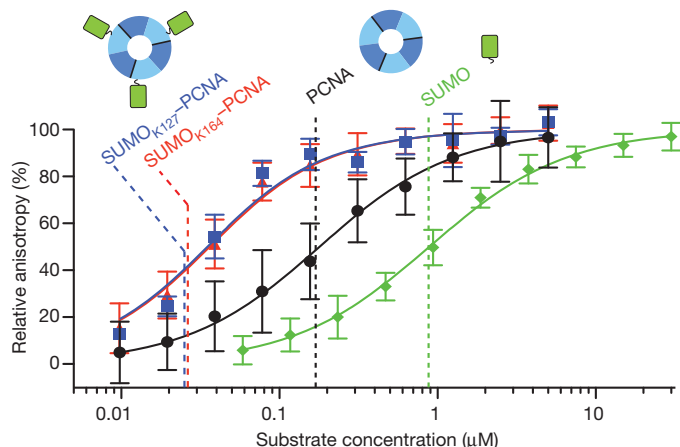


Figure 1 | Srs2 C-terminal domain interacts with PCNA, SUMO and SUMO-PCNA. Fluorescence anisotropy curves and calculated K_d values (dashed lines) for GST-SUMO (green), PCNA (black), SUMO_{K127}-PCNA (blue) and SUMO_{K164}-PCNA (red) titrations against 20 nM BODIPY-FL-Srs2_{1107–1174}. Cartoons of SUMO (green; right), PCNA (trimer, domain I (dark blue), domain II (light blue); middle) and SUMO-PCNA (left) above dashed lines. Assays conducted in triplicate. Error bars are ± 1 standard deviation; $n = 12$ for SUMO, $n = 12$ for PCNA, $n = 9$ for SUMO_{K164}-PCNA and $n = 3$ for SUMO_{K127}-PCNA. See Methods.

E3 ligase and mutations in PCNA (K127G; K164R) to direct SUMO modification to K164 or K127, respectively^{22,23}. K127 is conserved as glycine in most metazoan PCNA family members. Fluorescence polarization assays show that Srs2_{1107–1174} interacts with SUMO_{K164}-PCNA and SUMO_{K127}-PCNA with K_d values of 26.3 ± 3.1 nM and 25.1 ± 4.9 nM, respectively (Fig. 1, Supplementary Table 1 and Supplementary Figs 2 and 3), values sevenfold and 33-fold lower than for PCNA and GST-SUMO, respectively.

Structures of SUMO-conjugated PCNA

A structure of reconstituted SUMO_{K164}-PCNA was determined to a resolution of 2.6 Å (Fig. 2a, Supplementary Table 2 and Supplementary Fig. 5a). PCNA adopts the archetypal trimeric ring⁵ and electron density and the proximity of PCNA K164 and SUMO G98 are consistent with SUMO being conjugated to PCNA K164 (Supplementary Fig. 4a, c). The conformation observed for SUMO and its contacts to a PCNA loop between residues 184 and 198 (Supplementary Fig. 5a) are similar to another structure obtained by expressing PCNA as two self-assembling polypeptides split at residues 163 and 165 with SUMO introduced as an N-terminal fusion to the C-terminal peptide²⁴.

We found that *N*-ethylmaleimide (NEM) chemically modifies PCNA on cysteine side chains 22 and 81 (Supplementary Figs 4d and 6), causing PCNA to run on gel filtration at an apparent size that is three times smaller than the PCNA trimer (Supplementary Fig. 7a). We term this preparation PCNA^{mono}. The structure of SUMO_{K164}-PCNA^{mono} was determined at 2.8 Å resolution (Fig. 2b, Supplementary Table 2 and Supplementary Fig. 5a). Although a monomer on gel filtration, SUMO_{K164}-PCNA^{mono} crystallizes as a continuous right-handed helix with four PCNA monomers comprising a single turn (Fig. 2b) through formation of a PCNA-PCNA interface similar to that observed in SUMO_{K164}-PCNA^{tri} and PCNA^{tri} (ref. 5) (Supplementary Fig. 6). The right-handed helical configuration is in agreement with molecular dynamics simulations of dimeric PCNA²⁵, cryo-electron microscopy studies of the archaeal clamp-clamp loader-DNA complex²⁶ and recent studies in which the T4 clamp loader was crystallized with an open ring²⁷ (Supplementary Fig. 8). In this structure SUMO makes few contacts to PCNA and adopts a different orientation compared to SUMO_{K164}-PCNA^{tri} (Supplementary Fig. 5a). This indicates SUMO can achieve different conformations in solution, although this conformation is probably stabilized by lattice contacts (Supplementary Fig. 5b, c).

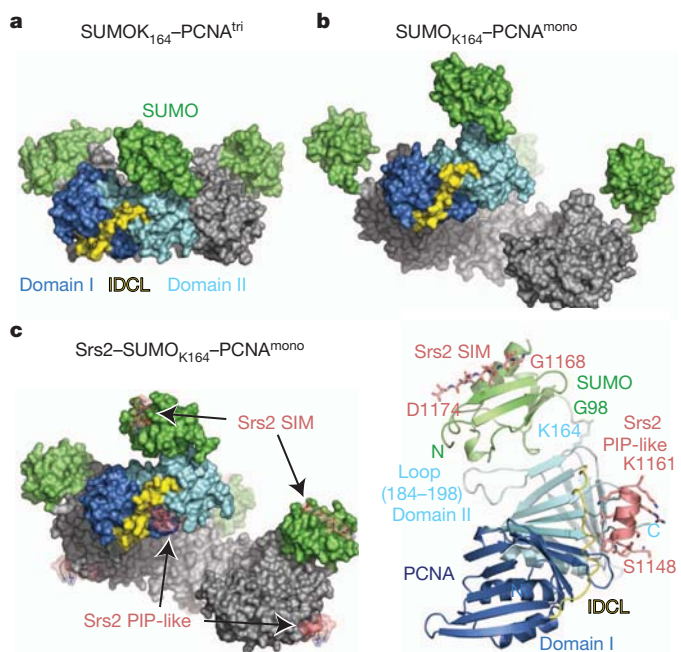


Figure 2 | Structures of SUMO-PCNA alone and in complex with Srs2.

a, Surface representation for SUMO_{K164}-PCNA^{tri}. PCNA domain I (dark blue), domain II (light blue), IDCL (yellow) and SUMO (green) are labelled. **b**, SUMO_{K164}-PCNA^{mono} as in **a**. Four protomers oriented to visualize the 'open ring'. **c**, Srs2_{1107–1174}-SUMO_{K164}-PCNA^{mono} as in **a** and **b** (left). Srs2 SIM and PIP-like motifs labelled (pink; stick and transparent surface). Cartoon representation for one protomer of Srs2_{1107–1174}-SUMO_{K164}-PCNA^{mono} (right) coloured as in **c**. PCNA N- and C- termini, loop 184–198, and K164 and SUMO N terminus and G98 are labelled. Srs2 SIM and PIP-like motifs labelled with single amino acid code and number to indicate the termini of each motif.

Structure of Srs2-SUMO-PCNA

Srs2_{1107–1174} interacts with SUMO_{K164}-PCNA^{mono} and SUMO_{K164}-PCNA^{tri} by analytical gel filtration (Supplementary Fig. 7b) and binds SUMO_{K164}-PCNA^{mono} and SUMO_{K127}-PCNA^{mono} with apparent K_d values of 79 ± 18 nM and 14.0 ± 3.9 nM, threefold worse and less than twofold better than SUMO_{K164}-PCNA^{tri} or SUMO_{K127}-PCNA^{tri}, respectively. Our data show that Srs2 interacts with NEM-treated PCNA^{mono} 6.5-fold worse than PCNA^{tri}. We posit that outside of its conformational constraints in the trimer PCNA^{mono} is more flexible and this probably results in conformational heterogeneity in the PIP-box binding site. In spite of this difference Srs2 does show specificity in binding SUMO-PCNA because the K_d of interactions with SUMO_{K164}-PCNA^{mono} and SUMO_{K127}-PCNA^{mono} is 14-fold and 78-fold lower than that of PCNA^{mono}, respectively (Supplementary Table 1), indicating that SUMO-PCNA^{mono} is a reasonable surrogate for SUMO-PCNA^{tri} in its interactions with Srs2. Furthermore, these data indicate that Srs2 recognizes SUMO-PCNA within a single SUMO-PCNA protomer and does not bridge protomers in the PCNA trimer. This hypothesis is further supported by distance constraints between the PIP-box binding site and lysine residues as intra-protomer distances to K127 or K164 are two to three times shorter than distances to K127 or K164 in adjacent protomers (see below).

Complexes between Srs2_{1107–1174} and SUMO_{K164}-PCNA^{tri} or SUMO_{K164}-PCNA^{mono} were purified (Supplementary Fig. 7b). We obtained crystals and determined the structure for Srs2-SUMO_{K164}-PCNA^{mono} at 2.9 Å resolution (Fig. 2c, Supplementary Table 2 and Supplementary Fig. 4c). In this structure, PCNA adopts a right-handed helical conformation similar to apo-SUMO_{K164}-PCNA^{mono} (PCNA protomers align to a root mean squared deviation (r.m.s.d.) of 0.98 Å) whereas SUMO adopts a conformation similar to SUMO_{K164}-PCNA^{tri} (Supplementary Fig. 5a). Electron density for Srs2 residues 1148–1161 is located in the interface between PCNA domains I, II

and the IDCL, and electron density for Srs2 residues 1168–1174 is adjacent to SUMO β 2 and α 1 (Fig. 2c and Supplementary Fig. 4b). Electron density for Srs2 residues 1107–1147 or 1162–1167 is not observed in our structure. To eliminate the contribution of residues N-terminal to the motifs defined in our structure to recognition of SUMO–PCNA, we prepared an Srs2 fragment encompassing residues 1137–1174. Although the interaction of Srs2_{1137–1174} with PCNA and SUMO individually is three- and fourfold weaker, respectively, than that of Srs2_{1107–1174}, the interaction of Srs2_{1137–1174} with SUMO–PCNA was eightfold stronger than that of Srs2_{1107–1174} (Supplementary Table 1 and Fig. 3).

Functional analysis of the Srs2 PIP-box

The Srs2 PCNA-interaction motif differs from canonical PIP-box motifs^{6–8,21,28,29}, perhaps consistent with its not being described previously. Most PIP-boxes contain a core Qxx Ψ motif followed by two conserved aromatic residues three and four residues C-terminal of the core motif (Fig. 3a and Supplementary Fig. 9). Srs2 residues 1149–1152 (QMDI) conform to a canonical PIP-box motif (Qxx Ψ), but Srs2 residues 1153–1161 lack the conserved aromatic residues. In addition, most PIP-boxes adopt a 3_{10} helix after the Qxx Ψ motif whereas Srs2 residues 1153–1161 adopt two turns of α -helix that project Q1155 and L1156 into the PCNA surface (Fig. 3b).

As the Srs2 PCNA-interaction motif differs from canonical PIP-box motifs, we conducted mutational and functional studies to evaluate the importance of side-chain contacts between Srs2 and PCNA. Q1149 is within hydrogen bonding distance of the backbone carbonyl oxygen of PCNA A251 (Fig. 3b), and Q1149E substitution results in a modest threefold defect in interaction with PCNA^{tri} (Fig. 4a). The D1151 side chain caps the N-terminal end of the Srs2 helix with a hydrogen bond to the backbone amide of S1154 (Fig. 3b). D1151A elicits a tenfold defect in binding PCNA^{tri} (Fig. 4a). The I1152 side chain projects into a hydrophobic pocket on PCNA and a main-chain hydrogen bond is observed between the I1152 backbone amide and the carbonyl oxygen of PCNA R44 (Fig. 3b). I1152A results in a 60-fold defect in interaction with PCNA^{tri} (Fig. 4a).

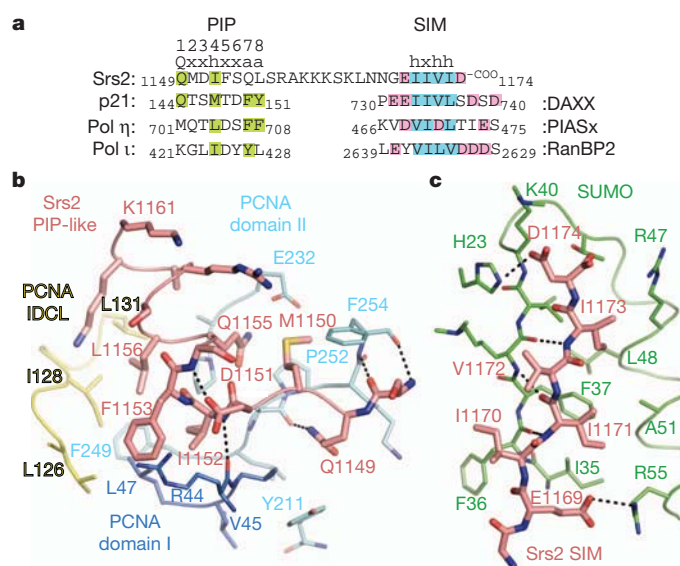


Figure 3 | Srs2 PIP-like motif and SIM interactions with PCNA and SUMO. **a**, Sequence alignment for Srs2 CTD (amino acids 1149–1174) above p21, Pol η and Pol ι PIP-box motifs and DAXX, PIASx and RanBP2 SIMs. PIP-box consensus motif numbered above Srs2 (Q, glutamine; x, any amino acid; h, hydrophobic; a, aromatic). Amino acids similar to consensus are highlighted (green). Hydrophobic and acidic residues highlighted blue and pink for SIMs. **b**, Srs2 PIP-like motif (pink), PCNA coloured as in Fig. 2. PCNA, Srs2 and select amino acids labelled by numbered single-letter code. Hydrogen bonds as dashed lines. **c**, Srs2 SIM (pink) and SUMO (green) as in **b**.

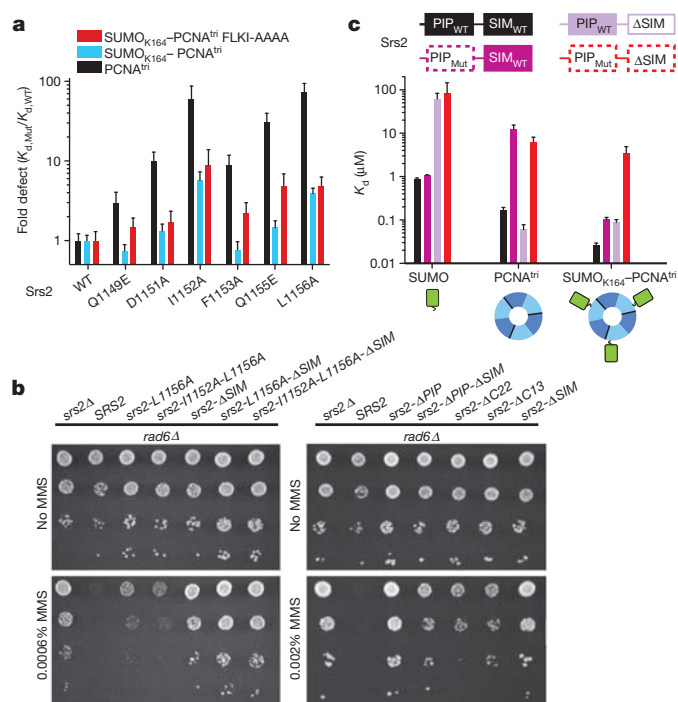


Figure 4 | Srs2 PIP-like motif and SIM required for recognition of SUMO–PCNA. **a**, Bar graphs showing fold-defect in K_d (y axis) for wild-type and Srs2_{1107–1174} mutants (x axis) for PCNA^{tri} K127G (black), SUMO_{K164}–PCNA^{tri} K127G (blue), or SUMO_{K164}–PCNA^{tri} FLKI_{125–128}AAAA (red). **b**, Suppression of *rad6Δ* DNA damage sensitivity by *srs2Δ* or *srs2* alleles. Shown are tenfold serial dilutions of cells on plates containing MMS as indicated (left). **c**, Bar graphs showing K_d values for interaction of Srs2 containing a wild-type PIP and SIM (PIP_{WT}–SIM_{WT}) (black), a mutant PIP-like motif (L1156A; PIP_{Mut}) and SIM_{WT} (dark purple; dashed line around PIP_{Mut}), a PIP_{WT} with SIM deletion (Δ SIM) (light purple; dashed line around Δ SIM), and a PIP_{Mut} Δ SIM (red dashed line) with SUMO (left), PCNA (middle) and SUMO_{K164}–PCNA^{tri} (right). Cartoons coloured as in Fig. 1. Fluorescence polarization assays conducted in triplicate. Error bars are ± 1 standard deviation. See Methods.

The α -helical geometry observed for Srs2 residues 1153–1161 projects side chains of F1153, Q1155 and L1156 into a mostly hydrophobic pocket on PCNA composed of residues V45 and L47 from domain I; L126, I128 and L131 from the IDCL; and P234, F249 and P252 from domain II (Fig. 3b). In addition, the Q1155 side chain Ne is within 4 Å of the PCNA E232 carboxylate. Although these Srs2 residues do not conform to side chains conserved in most other PIP-box motifs (Fig. 3a and Supplementary Fig. 9), each are important for interaction with PCNA because Srs2 containing F1153A, L1156A or Q1155E substitutions bind PCNA^{tri} with 9-, 73- or 31-fold reduced affinity relative to wild-type Srs2_{1107–1174}, respectively (Fig. 4a).

PCNA side chains proximal to the Srs2 PIP-like motif include L126 and I128 in the IDCL (Fig. 3b), and alanine substitutions in the PCNA IDCL (FLKI_{125–128}AAAA) weaken interaction with Srs2_{1107–1174} by more than 60-fold (Supplementary Table 1). Furthermore, defects observed for Srs2 Q1149E, F1153A and Q1155E are exacerbated in fluorescence polarization assays when interactions with SUMO_{K164}–PCNA^{tri} FLKI_{125–128}AAAA and SUMO_{K164}–PCNA^{tri} are compared (Fig. 4a). Conformations of PCNA IDCLs vary when Srs2_{1107–1174}–SUMO_{K164}–PCNA^{mono} and SUMO_{K164}–PCNA^{mono} are compared to other PCNA structures (Supplementary Fig. 10); it is possible that contacts to Srs2 elicit some of these differences.

Srs2 I1152A and L1156A substitutions result in the greatest defects in interaction with PCNA and SUMO_{K164}–PCNA^{tri} (Fig. 4a). To evaluate if these substitutions could impair Srs2 functions *in vivo* with respect to suppression of the DNA damage sensitivity of *rad6Δ* or *rad18Δ* strains, we constructed *rad6Δsrs2Δ* and *rad18Δsrs2Δ* and

complemented strains with *SRS2* or mutant *srs2* alleles under control of the endogenous *SRS2* promoter. Consistent with the importance of the PIP-like motif *in vitro*, strains harbouring *srs2-L1156A* or *srs2-L1152A-L1156A* were more resistant to the DNA damaging agent methyl methanesulphonate (MMS) than strains containing wild-type *SRS2*, although both were more sensitive to MMS than strains harbouring *srs2* with a deletion of the PIP-like motif (Fig. 4b and Supplementary Fig. 11).

Functional analysis of the Srs2 SIM

The Srs2 SIM (residues 1168–1174) forms a β -strand that interacts with SUMO β 2 and α 1 (Fig. 3c). The parallel orientation of the SIM and SUMO β 2 in our structure is distinct from SUMO complexes with SIMs of RanBP2 (ref. 30), thymine DNA glycosylase³¹ and the SUMO E1 (ref. 32) but similar to PIASx³³, DAXX²⁰ and MCAF1 (ref. 34) (Supplementary Fig. 12). Consistent with the structure and previous studies¹², removal of the SIM by deletion of Srs2 residues 1168–1174 has the same effect as removing SUMO from PCNA because Srs2_{1107–1167 Δ SIM} interaction with SUMO–PCNA is equivalent to PCNA alone (Fig. 4c and Supplementary Table 1).

Six residues separate the Srs2 PIP-like motif (1148–1161) and SIM (1168–1174), too few to span 39 Å between these motifs in our structure even if extended (~20 Å) (Fig. 2c). We posit that the Srs2 SIM exchanged interactions with another SUMO in the lattice during crystallization as the N termini of two symmetry-related SIMs are located within 15 Å of the PIP-like motif (Supplementary Fig. 13a). It is important to note that Srs2 interacts equally well with SUMO_{K164}–PCNA^{tri} and SUMO_{K127}–PCNA^{tri} in solution (see above). For Srs2 to simultaneously engage PCNA and SUMO in these complexes SUMO would have to adopt a different conformation from that observed in SUMO_{K164}–PCNA^{tri} or SUMO_{K164}–PCNA^{mono} structures.

Models were generated for SUMO_{K164}–PCNA^{tri} and SUMO_{K127}–PCNA^{tri} to permit simultaneous engagement of the Srs2 SIM and PIP elements with PCNA and SUMO in a single protomer (Fig. 5 and Supplementary Fig. 13b). These models suggest that SUMO would adopt conformations distinct from that observed in most SUMO–PCNA structures where SUMO interacts with PCNA loop 184–198. Consistent with our models, fluorescence polarization data show that V186D and MEH_{188–190}AAA mutations in PCNA loop 184–198, substitutions predicted to disrupt non-covalent contacts between SUMO and PCNA, have no impact on the ability of Srs2 to interact with SUMO_{K164}–PCNA (Supplementary Table 1).

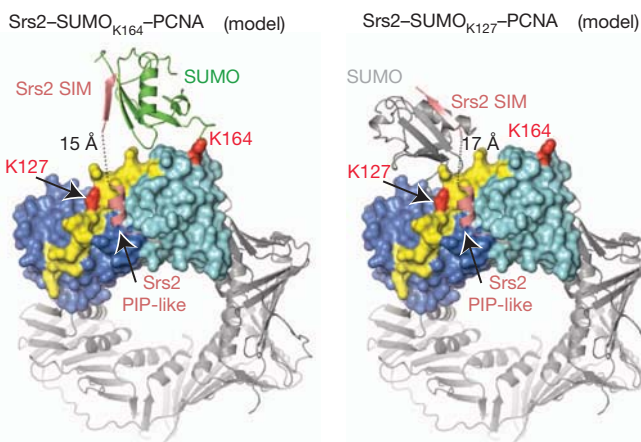


Figure 5 | Models for Srs2–SUMO–PCNA complexes. Models for SUMO conformations to enable simultaneous interaction with the SIM and PIP-like motif when attached to PCNA K164 (left) or K127 (right). Models generated required a simple rotation of SUMO at the isopeptide linkage for SUMO_{K164}–PCNA and a slight rotation and translation of SUMO from a symmetry related complex to mimic SUMO_{K127}–PCNA (Supplementary Fig. 13).

The α -helical conformation of the Srs2 PIP-like motif differs from other PIP-box motifs that make more extensive contacts to the IDCL (Supplementary Fig. 9). This conformation points the Srs2 C-terminal end of the PIP-like motif away from the PCNA surface towards SUMO in our SUMO–PCNA models (Fig. 5). Furthermore, previous studies suggested that SUMO modification of PCNA antagonizes Eco1, possibly by interfering with its ability to interact with PCNA via its PIP-box³⁵. We purified Eco1 and used it in pull-down assays to determine if Eco1 interacts with PCNA, SUMO_{K127}–PCNA or SUMO_{K164}–PCNA. Eco1 interacts with PCNA whereas PCNA FLK_{125–128}AAAA substitution diminished interaction (Supplementary Fig. 14). This is consistent with the IDCL mediating interactions with PIP-box proteins. We find that Eco1 binds SUMO_{K127}–PCNA or SUMO_{K164}–PCNA at levels similar to that observed for PCNA. Although it is clear that the SUMO pathway antagonizes PCNA-dependent Eco1 functions *in vivo*³⁵, our data exclude a simple steric occlusion model because SUMO modification on PCNA K127 and K164 does not prevent Eco1 from interacting with PCNA.

Srs2 recognition of SUMO–PCNA

Our data show that Srs2 requires the SIM to recognize SUMO and the PIP-like motif to recognize PCNA and that both elements are required to specifically recognize SUMO–PCNA. Consistent with this hypothesis, interaction with SUMO is strictly dependent on the SIM but independent of mutation in the PIP-like motif whereas interaction with PCNA is dependent on the integrity of the PIP-like motif and independent of the SIM (Fig. 4c). Finally, both PIP-like and SIM motifs are required to achieve specificity during recognition of SUMO–PCNA because mutation of the PIP-like motif or deletion of the SIM lessen interaction whereas mutations in both prevent interaction with SUMO–PCNA (Fig. 4c).

Although mutation or deletion of individual motifs results in Srs2 variants that interact with SUMO or PCNA in SUMO–PCNA *in vitro*, it is important to note that both mutants lack elements required to specifically recognize SUMO–PCNA. To validate this hypothesis for Srs2 function *in vivo*, we complemented *rad6 Δ srs2 Δ* and *rad18 Δ srs2 Δ* with Srs2 variants lacking the PIP-like motif, the SIM motif or both. As predicted, these variants behave similarly to a strain lacking *SRS2* as each suppresses the DNA-damage sensitivity of *rad6 Δ* and *rad18 Δ* strains (Fig. 4b and Supplementary Fig. 11). Importantly, strains harbouring a SIM deletion in conjunction with single- and double-point substitutions in the PIP-like motif are more resistant to DNA damage than strains containing full-length Srs2 with mutations in the PIP-like motif.

Discussion

Recognition of Ub- and Ubl-modified substrates is a critical first step in Ub- and Ubl-mediated signal transduction. Details of this process have been widely assumed³⁶, but the molecular bases for these interactions have not been formally demonstrated. Here we identify one mechanism to achieve specificity during recognition of SUMO-modified PCNA. Given the modular nature of PIP-box motifs and ubiquitin-binding domains in some translesion polymerase family members^{21,37}, this mechanism may apply to their recognition of Ub–PCNA³⁸. It is also worth noting that a recent study claims to have identified a mammalian Srs2 orthologue named PARI which contains a C-terminal canonical PIP-box and an internal SIM³⁹, although these motifs differ in comparison to Srs2 with respect to their position in primary sequence.

It seems likely that the mechanism employed by Srs2 to recognize SUMO–PCNA, namely juxtaposition of domains or motifs in the receptor that simultaneously engage the substrate and Ub and Ubl modifier within the context of the Ub- and Ubl-conjugated substrate, will be used in other signal transduction pathways that rely on receptors to specifically recognize Ub- and Ubl-modified substrates.

METHODS SUMMARY

Yeast proteins were expressed in *Escherichia coli* and purified. PCNA was conjugated to SUMO using the E3 ligase Siz1. For fluorescence polarization, substrates were titrated in triplicate against 20 nM BODIPY-FL-conjugated Srs2 and data fit to a single-site binding model accounting for ligand depletion. Crystals were obtained by vapour diffusion. Diffraction data were collected at beam lines X29 (NSLS) and 24-IDC (APS), phases were calculated by molecular replacement. Yeast strains were constructed and complemented with plasmids containing SRS2 under the control of its endogenous promoter.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 August 2011; accepted 23 January 2012.

- Kirkin, V. & Dikic, I. Role of ubiquitin- and Ubl-binding proteins in cell signaling. *Curr. Opin. Cell Biol.* **19**, 199–205 (2007).
- Gareau, J. R. & Lima, C. D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nature Rev. Mol. Cell Biol.* **11**, 861–871 (2010).
- Dikic, I., Wakatsuki, S. & Walters, K. J. Ubiquitin-binding domains – from structures to functions. *Nature Rev. Mol. Cell Biol.* **10**, 659–671 (2009).
- Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **129**, 665–679 (2007).
- Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M. & Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243 (1994).
- Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* **87**, 297–306 (1996).
- Bruning, J. B. & Shamoo, Y. Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase- δ p66 subunit and flap endonuclease-1. *Structure* **12**, 2209–2219 (2004).
- Vijayakumar, S. *et al.* The C-terminal domain of yeast PCNA is required for physical and functional interactions with Cdc9 DNA ligase. *Nucleic Acids Res.* **35**, 1624–1637 (2007).
- Scott, M. T., Morrice, N. & Ball, K. L. Reversible phosphorylation at the C-terminal regulatory domain of p21(Waf1/Cip1) modulates proliferating cell nuclear antigen binding. *J. Biol. Chem.* **275**, 11529–11537 (2000).
- Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. & Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135–141 (2002).
- Stelter, P. & Ulrich, H. D. Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature* **425**, 188–191 (2003).
- Pfander, B., Moldovan, G. L., Sacher, M., Hoege, C. & Jentsch, S. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **436**, 428–433 (2005).
- Papouli, E. *et al.* Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol. Cell* **19**, 123–133 (2005).
- Krejci, L. *et al.* DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
- Veaute, X. *et al.* The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
- Lawrence, C. W. & Christensen, R. B. Metabolic suppressors of trimethoprim and ultraviolet light sensitivities of *Saccharomyces cerevisiae* rad6 mutants. *J. Bacteriol.* **139**, 866–876 (1979).
- Baba, D. *et al.* Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**, 979–982 (2005).
- Ulrich, H. D. PCNASUMO and Srs2: a model SUMO substrate-effector pair. *Biochem. Soc. Trans.* **35**, 1385–1388 (2007).
- Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G. & Chen, Y. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc. Natl Acad. Sci. USA* **101**, 14373–14378 (2004).
- Chang, C. C. *et al.* Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. *Mol. Cell* **42**, 62–74 (2011).
- Hishiki, A. *et al.* Structural basis for novel interactions between human translesion synthesis polymerases and proliferating cell nuclear antigen. *J. Biol. Chem.* **284**, 10552–10560 (2009).
- Yunus, A. A. & Lima, C. D. Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. *Mol. Cell* **35**, 669–682 (2009).
- Yunus, A. A. & Lima, C. D. Purification of SUMO conjugating enzymes and kinetic analysis of substrate conjugation. *Methods Mol. Biol.* **497**, 167–186 (2009).
- Freudenthal, B. D., Brogie, J. E., Gakhar, L., Kondratik, C. M. & Washington, M. T. Crystal structure of SUMO-modified proliferating cell nuclear antigen. *J. Mol. Biol.* **406**, 9–17 (2011).
- Kazmirski, S. L., Zhao, Y., Bowman, G. D., O'Donnell, M. & Kuriyan, J. Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. *Proc. Natl Acad. Sci. USA* **102**, 13801–13806 (2005).
- Miyata, T. *et al.* Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc. Natl Acad. Sci. USA* **102**, 13795–13800 (2005).
- Kelch, B. A., Makino, D. L., O'Donnell, M. & Kuriyan, J. How a DNA polymerase clamp loader opens a sliding clamp. *Science* **334**, 1675–1680 (2011).
- Bowman, G. D., O'Donnell, M. & Kuriyan, J. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* **429**, 724–730 (2004).
- Sakurai, S. *et al.* Structural basis for recruitment of human flap endonuclease 1 to PCNA. *EMBO J.* **24**, 683–693 (2005).
- Reverter, D. & Lima, C. D. Insights into E3 ligase activity revealed by a SUMO–RanGAP1–Ubc9–Nup358 complex. *Nature* **435**, 687–692 (2005).
- Baba, D. *et al.* Crystal structure of SUMO-3-modified thymine-DNA glycosylase. *J. Mol. Biol.* **359**, 137–147 (2006).
- Olsen, S. K., Capili, A. D., Lu, X., Tan, D. S. & Lima, C. D. Active site remodelling accompanies thioester bond formation in the SUMO E1. *Nature* **463**, 906–912 (2010).
- Song, J., Zhang, Z., Hu, W. & Chen, Y. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J. Biol. Chem.* **280**, 40122–40129 (2005).
- Sekiya, N. *et al.* Structure of the small ubiquitin-like modifier (SUMO)-interacting motif of MBD1-containing chromatin-associated factor 1 bound to SUMO-3. *J. Biol. Chem.* **283**, 35966–35975 (2008).
- Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA controls establishment of sister chromatid cohesion during S phase. *Mol. Cell* **23**, 723–732 (2006).
- Seet, B. T., Dikic, I., Zhou, M. M. & Pawson, T. Reading protein modifications with interaction domains. *Nature Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
- Bienko, M. *et al.* Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821–1824 (2005).
- Chen, J., Ai, Y., Wang, J., Haracska, L. & Zhuang, Z. Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis. *Nature Chem. Biol.* **6**, 270–272 (2010).
- Moldovan, G. L. *et al.* Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Mol. Cell* **45**, 75–86 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Kuriyan and B. Kelch for coordinates of the T4 clamp before publication. NE-CAT beamlines (Advanced Photon Source) supported by RR-15301 (NIH NCRR). APS supported by the US Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Beamline X29 (National Synchrotron Light Source) supported by the US Department of Energy, the Office of Basic Energy Sciences and P41RR012408 (NIH NCRR). A.A.A., F.M. and C.D.L. are supported by NIH R01 GM065872 to C.D.L. and F32 GM086066 to A.A.A.

Author Contributions Experiments performed and analysed by A.A.A., F.M. and C.D.L. Manuscript prepared by A.A.A. and C.D.L.

Author Information Atomic coordinates and structure factors are deposited in PDB under accession codes 3V60, 3V61 and 3V62. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.D.L. (limac@mskcc.org).

METHODS

Cloning, expression and purification of recombinant proteins. Expression and purification of yeast E1 (Aos1/Uba2), E2 (Ubc9), yeast SUMO (NA18 Smt3), Siz1_{112–465} and affinity-tag-free yeast PCNA (wild type, K127G and FLKI_{125–128}AAAA) have been described^{22,23}. PCNA K164R and SUMO K19R were generated by PCR-based site-directed mutagenesis. Srs2_{1027–1174} was obtained by PCR from *S. cerevisiae* genomic DNA and cloned into the pSMT3 vector encoding an N-terminal His₆-Smt3 tag. Fragments, including the Srs2_{1107–1174} construct, and mutants were generated by subcloning and by PCR-based site-directed mutagenesis. Srs2 was expressed in BL21 (DE3) RIL Codon Plus cells after induction with 0.5 mM isopropylthiogalactoside (IPTG) and incubation of the cultures at 30 °C for 3 h. Cell pellets were suspended in 2.5× volume of 20% sucrose, 50 mM Tris, pH 8.0 and lysed by sonication. Srs2 was isolated from cleared lysate by immobilized metal-affinity chromatography using Ni-NTA resin (Qiagen). Srs2 was eluted from the resin in buffer containing 250 mM imidazole, 350 mM NaCl, 20 mM Tris, pH 8.0, 1 mM β-mercaptoethanol (βME), and the His₆-Smt3 tag was subsequently cleaved by addition of Ulp1 to a 1:1,000 mass ratio followed by overnight incubation/dialysis into 350 mM NaCl, 20 mM Tris, pH 8.0, 1 mM βME at 4 °C (ref. 40). Ulp1 was inactivated by adding NEM to a final concentration of 5 mM followed by incubation at room temperature for 30 min and 4 °C for 1 h. Srs2 was separated from other proteins in this mixture by size-exclusion chromatography (Superdex75 26/60; GE Healthcare; equilibrated with 350 mM NaCl, 20 mM Tris, pH 8.0, 1 mM βME). Fractions containing Srs2 were combined, dialysed overnight against 50 mM NaCl, 20 mM Tris, pH 8.0, 1 mM βME and purified by anion-exchange chromatography (MonoQ 10/10; GE Healthcare). Srs2_{1107–1174} was recovered in the unbound fraction, and Srs2_{1027–1174} was eluted using a gradient from 50 mM to 475 mM NaCl over 12 column volumes in 20 mM Tris, pH 8.0, 1 mM βME. Srs2 was concentrated using an Amicon Ultra centrifugal filter, snap frozen in liquid nitrogen and stored at –80 °C. Eco1 was cloned from *S. cerevisiae* genomic DNA into the pET28a vector encoding an N-terminal hexahistidine tag. Eco1 was expressed in BL21 (DE3) RIL Codon Plus cells and purified through IMAC steps according to the same protocols described for expression and purification of Srs2 fragments. Upon elution from the Ni-NTA resin, Eco1 was purified by size-exclusion chromatography (Superdex75 26/60; GE Healthcare; equilibrated with 350 mM NaCl, 20 mM Tris, pH 8.0, 1 mM βME). Peak fractions were combined and Eco1 was concentrated to 8 mg ml^{–1} before snap freezing in liquid nitrogen for storage at –80 °C.

Reductive methylation of SUMO and preparation of PCNA^{mono}. Reductive methylation of yeast SUMO (NA18 K19R Smt3; *SUMO where the asterisk indicates reductively methylated) used paraformaldehyde and sodium borohydride as the methyl donor and reducing agent, respectively⁴¹. Monomeric PCNA or monomeric SUMO-PCNA was prepared by adding *N*-ethylmaleimide (NEM) to a final concentration of 5 mM to the corresponding trimeric species, incubating for 2 h at 4 °C, and then purifying by size exclusion chromatography (Superdex200 26/60).

Reconstitution of SUMO-PCNA. Yeast SUMO was conjugated to yeast PCNA using the SUMO E3 ligase Siz1 by mixing 15 μM PCNA, 65 μM SUMO (NA18 K19R Smt3 or reductively methylated NA18 K19R Smt3), 90 nM Aosl-His₆-Uba2 ΔCT (E1), 300 nM Ubc9 (E2), and 2 μM Siz1_{112–465} (E3) and 2 mM ATP, incubated at 37 °C for 2 h, and stopped by chromatographic resolution of the reaction components.

Fluorescence polarization assays. A cysteine residue was added N-terminal to Srs2_{1107–1174} which itself lacked any cysteine residues. BODIPY FL (Invitrogen) was conjugated to the thiol group via maleimide chemistry. For fluorescence polarization assays, substrate was titrated against a fixed concentration (20 nM) of labelled Srs2_{1107–1174} by mixing equal volumes of Srs2 peptide and serially diluted protein. Binding reaction conditions were 50 mM NaCl, 20 mM Tris, pH 8.0, 4 mM dithiothreitol (DTT). 20-μl reaction volumes were added to the wells of a 384-well microplate, and measurements taken using a SpectraMax M5 microplate reader (Molecular Devices) using an excitation wavelength of 485 nm, emission wavelength of 538 nm, and a cutoff of 530 nm. Experiments were performed in triplicate. Using the software package Prism 5 (GraphPad Software) data were fit to a single site binding model accounting for ligand depletion:

$$A = A_f + (A_b - A_f) \left(\frac{([Srs2] + K_d + [Prot]) - \sqrt{([Srs2] + K_d + [Prot])^2 - 4[Srs2][Prot]}}{2[Srs2]} \right)$$

where *A* is the measured anisotropy, [Srs2] is the fixed concentration of labelled receptor, and [Prot] is the total concentration of substrate. *A_b* and *A_f* are limiting anisotropies for bound and free peptide, respectively, and *K_d* is the dissociation constant.

For displacement assays, BODIPY-labelled Srs2_{1107–1174} at 40 nM was pre-incubated with PCNA^{tri} K127G, GST-SUMO or SUMO_{K164}-PCNA^{tri} at 600,

2,000 and 80 nM, respectively. Equal volumes of the pre-incubated binding reaction and serially diluted unlabelled Srs2_{1107–1174} were then mixed. 20-μl reaction volumes were added to the wells of a 384-well microplate, and measurements were made as for the direct fluorescence polarization assays. Experiments were performed in triplicate, and the data were fit using Prism 5 to a one-site binding model for competition assays.

Assessment of protein-protein interactions by analytical gel filtration. Purified proteins were mixed at the indicated concentrations in a volume of 220 μl and centrifuged, and 200 μl was loaded onto a Superose12 HR 10/30 column (GE Healthcare) equilibrated in 50 mM NaCl, 20 mM Tris, pH 8.0, 1 mM βME. The column was eluted at a flow rate of 0.5 ml min^{–1}, and 0.5-ml fractions were collected, analysed by SDS-PAGE and stained with Sypro Ruby Protein Gel Stain (Bio-Rad).

Assessment of protein-protein interactions by pull-downs. 400 μl of 750 μg ml^{–1} His₆-Eco1 in binding buffer (200 mM NaCl, 20 mM Tris, pH 8.0, 40 mM imidazole) was added to 50 μl of 50% (v/v) Ni-NTA resin (Qiagen) equilibrated with the same buffer and mixed for 20 min at 4 °C. Beads were subsequently washed with 3 × 400 μl of binding buffer to remove unbound His₆-Eco1. 400 μl of 10 μM prey protein in binding buffer was added to Eco1-coated beads and mixed for 15 min at 4 °C. Unbound prey protein was removed with 3 × 400 μl washes with binding buffer. Material was eluted from beads with successive 400 μl washes of binding buffer that contained increasing concentrations of NaCl. Eco1 and any remaining bound prey protein was eluted in a final wash with 100 μl 350 mM NaCl, 20 mM Tris, pH 8.0, 400 mM imidazole.

Crystallization and structure determination. Crystals were grown using the hanging drop vapour diffusion method by mixing equal volumes of sample and reservoir solution.

SUMO_{K164}-PCNA^{tri} K127G: purified SUMO_{K164}-PCNA^{tri} was crystallized at 18 °C in 4% polyethylene glycol (PEG) 8000, 500 mM LiSO₄ (pH 4.5). Crystals were transferred to cryoprotectant (reservoir supplemented with 15% ethylene glycol) before flash-freezing in liquid N₂. Diffraction data were collected at beamline 24-ID-C at the Advanced Photon Source at a wavelength of 0.979 Å. Crystals belong to space group F432 with one protomer of SUMO-PCNA in the asymmetric unit.

SUMO_{K164}-PCNA^{mono} K127G: Srs2_{1027–1174} was added to 8 mg ml^{–1} SUMO_{K164}-PCNA in threefold molar excess and dialysed against 100 mM NaCl, 20 mM Tris pH 8.0, 1 mM βME at 4 °C for 16 h. The Srs2_{1027–1174}:SUMO_{K164}-PCNA complex was purified on a Superdex 200 column and concentrated to ~8 mg ml^{–1}. Crystals were obtained at 4 °C in 21% MPD, 100 mM BaCl₂, 100 mM Bis-Tris, pH 6.5 and transferred to cryoprotectant (reservoir supplemented with 12% glycerol) before flash-freezing in liquid N₂. Diffraction data were collected at beamline 24-ID-C at the Advanced Photon Source at a wavelength of 0.979 Å. Crystals belong to space group I4₁ with one protomer of SUMO_{K164}-PCNA^{mono} in the asymmetric unit. It was subsequently determined that these crystals lacked the Srs2_{1027–1174} protein component, and that contaminating NEM, presumably present in earlier Srs2_{1027–1174} preparations, had modified PCNA C22 and C81.

SUMO_{K164}-PCNA K127G-Srs2_{1107–1174}: Srs2_{1107–1174} was added in threefold molar excess to SUMO_{K164}-PCNA^{tri} at a protomer concentration of 60 μM. NEM was added to a final concentration of 5 mM, and the sample incubated for 3 h at 4 °C before copurification of the complex by gel filtration (Superdex200 26/60; GE Healthcare; equilibrated with 100 mM NaCl, 20 mM Tris, pH 8.0, 1 mM βME). The complex was concentrated to ~10 mg ml^{–1} and crystallized at 6 °C in 1.9 M ammonium sulphate, 4% PEG400, 100 mM HEPES, pH 7.5 by the sitting drop vapour diffusion method. Crystals were washed and cryoprotected in 2 M ammonium sulphate, 4% PEG400, 100 mM HEPES, pH 7.5, 20% glycerol. Diffraction data were collected at beamline X29 of the National Synchrotron Light Source at a wavelength of 1.075 Å. Crystals belong to space group C2 with two Srs2-SUMO_{K164}-PCNA complexes in the asymmetric unit.

Crystallographic data were indexed, integrated and scaled using HKL2000 (ref. 42) and data reduced using programs in CCP4 (ref. 43). Data obtained for SUMO_{K164}-PCNA^{mono} were scaled maintaining separation of anomalous pairs to maintain the anomalous signal from the associated barium ions. The programs MOLREP⁴⁴ and PHASER⁴⁵ were used to find molecular replacement solutions using the coordinates of a yeast PCNA protomer from PDB 1PLQ and yeast SUMO from PDB 1EUV (Supplementary Table 2 and Methods). Models were manually inspected and rebuilt before refinement using the program O⁴⁶ and refined using CCP4's REFMAC⁴⁷. Models have reasonable geometry at the respective resolution as assessed by MolProbity⁴⁸. SUMO_{K164}-PCNA^{tri} has 95.5% and 3.9% of modelled residues in favoured and allowed Ramachandran regions, respectively with 0.6% outliers. SUMO_{K164}-PCNA^{mono} has 95.7% of modelled residues in favoured regions, 3.4% in allowed regions and 0.9% as outliers, and Srs2_{1107–1174}-SUMO_{K164}-PCNA^{mono} has 91.5% in favoured regions, 7.6% in

allowed regions and 0.9% as outliers. Structures of SUMO_{K164}-PCNA^{tri}, SUMO_{K164}-PCNA^{mono} and Srs2₁₁₀₇₋₁₁₇₄-SUMO_{K164}-PCNA^{mono} have Clash Scores in the 93rd, 95th and 87th percentiles, respectively, and MolProbity Scores in the 70th, 90th and 50th percentiles, respectively. All graphical representations of structure were generated using PYMOL⁴⁹.

Yeast complementation and growth assays. Yeast strains used in this study include *srs2Δrad6Δ* (MATa, *his3Δ1*, *leu2Δ0*, *lys2Δ0*, *ura3Δ0*, *srs2Δ::kanMX*, *rad6Δ::kanMX*) and *srs2Δrad18Δ* (MATa, *his3Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*, *srs2Δ::kanMX*, *rad18Δ::kanMX*). Double-deletion strains were constructed by mating *srs2Δ* (MATa *his3Δ1* *leu2Δ0* *lys2Δ0* *ura3Δ0*) to *rad6Δ* or *rad18Δ* (MATa *his3Δ1* *leu2Δ0* *met15Δ0* *ura3Δ0*) knockout strains from the *Saccharomyces* Genome Deletion Project (Open Biosystems). Diploids were selected on media lacking lysine and methionine then sporulated in 2% potassium acetate. The desired double-deletion strains were selected by screening spores by PCR to establish the genotype and by replica-plating spore colonies on selective media. Double-deletion strains were transformed with the pRS416 plasmid (URA3) or the pRS416 plasmid (URA3) containing SRS2 or *srs2* alleles flanked by 500 bp of its endogenous 5' and 3' UTRs. These constructs were generated by cloning the SRS2 5' UTR along with the SRS2 coding sequence corresponding to the first 1,115 residues into the XhoI and SalI sites of the pRS416 vector and the SRS2 3' untranslated region (UTR) into the XmaI and BamHI sites. Use of the SalI site introduced a DNA mutation that is silent with respect to the encoded Srs2 amino acid. The remainder of the SRS2 coding sequence was placed between the SalI and XmaI sites to produce the desired alleles including SRS2, *srs2-ΔPIPΔSIM* (Srs2₁₋₁₁₄₈), *srs2-ΔC22* (Srs2₁₋₁₁₅₂), *srs2-ΔC13* (Srs2₁₋₁₁₆₁), and *srs2-ΔSIM* (Srs2₁₋₁₁₆₇). *srs2-ΔPIP* (Srs2₁₋₁₁₄₈+PG+1168-1174) was generated by placing the SRS2 coding region corresponding to the C-terminal seven residues along with the 3' UTR into XmaI and BamHI sites. The SRS2 coding sequence corresponding

to residues 1115-1148 was then placed between the SalI and XmaI sites. This strategy introduced two non-native amino acids (Pro-Gly) between Srs2 residues 1148 and 1168. Yeast transformants were selected on media lacking uracil. For spotting assays to assess growth defects and drug sensitivity, cultures were grown overnight and diluted to a *D*₆₀₀ of 1.0. tenfold serial dilutions were made and 3 μl aliquots were spotted onto SD – Ura plates or SD – Ura + MMS plates containing the indicated concentration of DNA damage agent. Plates were incubated at 30 °C for ~3 days with photographs taken every 6–8 h.

40. Mossessova, E. & Lima, C. D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5**, 865–876 (2000).
41. Rayment, I. Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol.* **276**, 171–179 (1997).
42. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* vol. 276 (eds Carter, C. W. Jr. & Sweet, R. M.) 307–326 (Academic Press, 1997).
43. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Vagin, A. & Teplyakov, A. MOLREP: an automated program from molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
45. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
47. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
48. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
49. Delano, W. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).

CAREERS

UNITED STATES Competitors could benefit if grant applications go public **p.115**

ENVIRONMENT Research programme launches at US ecological sites **p.115**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

Z. FAULKES/SHUTTERSTOCK



A winning poster uses bold graphics, plenty of white space and limited text to draw viewers.

PRESENTATIONS

Billboard science

Posters are a chance to show off work and to network with colleagues, but only if the design is easy on the eye.

BY KENDALL POWELL

It was the scientific version of a Las Vegas-style casino. The poster session at the American Society for Cell Biology (ASCB) meeting in Denver, Colorado, in December 2011 included numerous aisles lined with posters, and young presenters showing off the fruits of endless hours in the laboratory. A few metres away, dozens of representatives from science agencies, institutes and scientific-instrument companies staffed a sea of exhibitor booths. They were shamelessly vying for

scientists' attention with the free sweets, pens, noise-makers and other bits of sparkly plastic that end up at the bottom of a canvas conference bag. It is difficult enough for an attendee to locate a particular poster; it is harder still for presenters to get noticed.

Rashieda Hatcher, a doctoral student at Baylor College of Medicine in Houston, Texas, and third-place winner in the graduate-student category of the ASCB Minority Affairs Committee poster contest, made the most of her prime, end-aisle location. Her display, which chronicled her study of the protein securin in

breast cancer, had many of the hallmarks of a winning poster: tastefully colourful, with an enlarged photo of her most important result (an abnormal branching of the mammary gland in mice that don't produce securin) at the centre. It had cartoons instead of text to explain the methods and conclusions. Hatcher had included business cards beside her poster so that viewers could contact her later. She had embraced the idea that a poster is a good place to sell her science.

Her efforts brought immediate results, with a visit from a leading researcher in her field. After talking to Hatcher about her work and future experiments, Zena Werb, a developmental biologist at the University of California, San Francisco, invited Hatcher to a talk by Werb's postdoctoral fellow. The next morning, they discussed the prospect of working together after Hatcher finishes her PhD.

Senior scientists see poster sessions as an opportunity for junior researchers to show off their science, to get fresh perspectives on research questions and to network. Presenters who follow a few basic rules in poster layout, and who talk viewers through their work, will draw a crowd. Judges say that a killer poster will have clean lines, white space, intriguing images and a clear visual flow that supports a well-told research story. The worst posters have panel after panel of tiny print, which can turn a poster session into a lonely four hours for the presenter.

"A poster should look catchy from 10 metres away," says Colin Purrington, formerly an evolutionary biologist at Swarthmore College in Pennsylvania, and now a photographer who maintains a blog on poster design (see tinyurl.com/6wb62m9). "A clean layout speaks to an uncluttered mind, and that's who you want to talk to at meetings," he says.

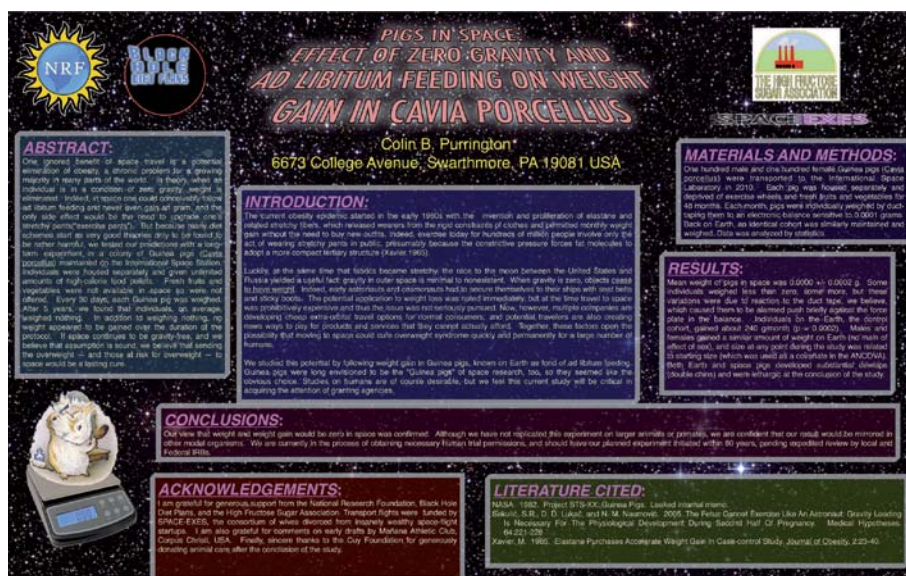
VISUAL AIDS

A poster should function as a visual aid — to help the presenter to talk through the research in an informal and engaging way. It should not be an abridged version of the researcher's current manuscript, stresses Zen Faulkes, a neuroethologist at University of Texas-Pan American in Edinburg. "People in a poster session usually have just come from an intense session and they need a mental break," says Jamie Simon, senior illustrator at the Salk Institute for Biological Sciences in La Jolla, California. Simon, who helps Salk researchers to design and print their posters, encourages scientists to stick to the standards of graphic design ►

► (see 'Creating an eye-catching poster'). The design should follow the photography law of thirds. Most posters are three columns across, with three sections per column. "Put the hot stuff in the middle of the second column," says Simon — as Hatcher did. It is more visually appealing to use lots of 'white space' to separate sections, and avoid distracting, whole-background images or swirls. Although most academic departments do not have a graphics guru like Simon, researchers can learn from poster sites such as Purrington's and Faulkes' blogs (see tinyurl.com/7uv92fu).

Giving a little thought to how the layout is organized goes a long way, says cell biologist Peter Lenart at the European Molecular Biology Laboratory in Heidelberg, Germany. Although most scientists use Microsoft PowerPoint software to design posters, Lenart prefers Adobe Illustrator because it offers greater flexibility and more graphic-design tools. He uses boxes to delineate each results section, each of which has its own headline, and places figure legends in boxes with arrows to direct readers to the appropriate image. Lenart places any additional information for that results panel in a box marked with an encircled 'i'. Clear direction gives even the most casual of browsers a chance to understand the overall point of the research.

Award-winning presenters weave a research tale that flows from the layout. Martin Bergert, a doctoral student at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, had two related stories he wanted to convey at the annual meeting of the European Molecular Biology Organization (EMBO) last September in Vienna. He placed the figures and text for each story in a box with a catchy title: "On the edge — blebs vs.



Colin Purrington's example of a 'bad' poster is crammed with text and has a distracting background.

lamellipodia" to highlight work on two types of cellular protrusion, and "On the move — how to migrate by blebbing" to describe results on cell migration.

Natasha Gutierrez, a doctoral student in cell biology at Rutgers University in Newark, New Jersey, and the second-place winner in the graduate-student category of the ASCB's poster competition, put all her potential figures onto a board, then removed those that didn't fit her research story. Gutierrez displayed her most impressive experiment in a three-panelled figure showing two microscopic views of her cells and a graph of quantitative results. She placed this figure in the middle, providing a focus for her oral presentation.

A well-organized poster alone is not enough to entice passers-by. Bergert's poster, which earned him a prize at the 2011 EMBO poster competition, had a 'hook' image to reel people in. He repeated a simple, unobtrusive image of a blebbing cell, the main focus of his poster, several times beneath the title. The images were visible from across the room and lured in scientists already interested in blebs and other people intrigued by the blobby cells.

Hook images also provide a visual entry into the poster's research topic. Purrington says that every poster should include a hook on the left-hand side to attract visitors who are unfamiliar with the work and will view it from a fresh perspective. The image could be a photograph of the research organism, a map or a portrait of a mathematician, for example. It is worth taking the time to come up with a creative hook, says Purrington, who notes that the largest crowd at a Swarthmore College poster session was at a chemist's poster that featured a photo of a man suffering from arsenic poisoning.

Presenters should make sure their figures are easy to interpret from a metre or so away. Axes and labels that have been taken from a previous talk or publication should be adjusted so that they are easily visible. Arrows and artificial colours can also be used to highlight a point. "Don't be afraid to write on top of a cell, 'This is the typical phenotype'. Make it as visually clear as possible," says Lenart.

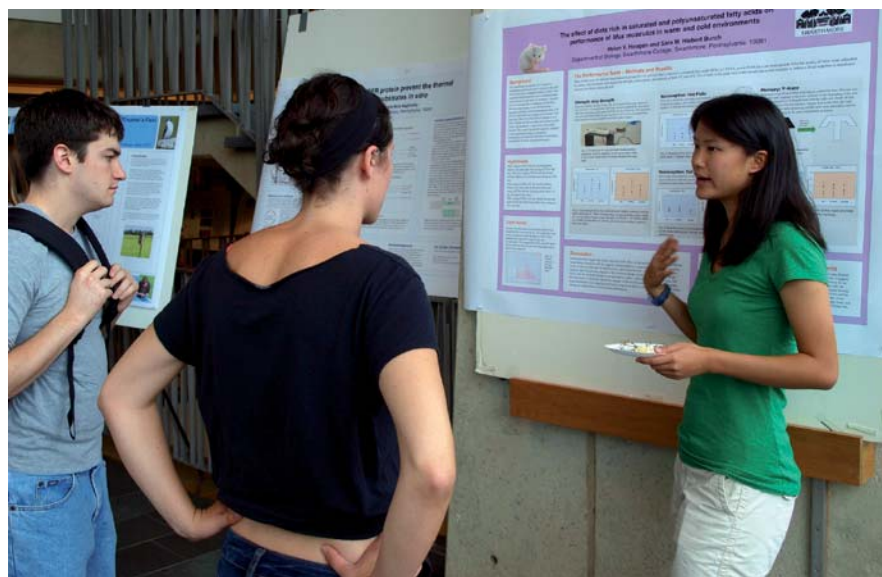
For simplicity, only the most important data points should be included in a poster, says Renato Aguilera, a judge for the ASCB contest and a molecular-cell biologist at the University of Texas at El Paso. "Have just two lines on your graph: the control and the most important result," he says, adding that posters should not force viewers to wade through hundreds of data points.

Finally, a few props can help to engage viewers. Bergert showed cell-migration movies on

CREATING AN EYE-CATCHING POSTER

Simple guidelines for poster presentation

- Choose landscape (horizontal) over portrait (vertical) orientations.
- Follow the recommended dimensions and instructions. Make sure you allow for white space and large fonts and images.
- Use photographs, cartoons or illustrations to explain concepts. Limit the word count to 1,000 words.
- Take great care in writing the abstract: conference attendees will identify posters to visit from the abstract book.
- Make the title the punchline of the research — and make it intriguing. Consider placing an engaging image close by.
- Titles and headings should be in a sans-serif font, such as Helvetica. Other text should be in a serif font such as Times New Roman, with a minimum size of 22 points.
- Consider short bullet points for methods and conclusions.
- Use black text on a white background. Red text can be used to draw attention, but avoid blue and yellow, which are hard to read.
- Place figures in an obvious order, and consider using numbering. Figures should have a large headline with the main finding.
- Enlarge the best piece of data and place it squarely in the middle at eye level.
- Have someone else proofread the text.
- Check the poster on a large computer screen at 100%, then step back half a metre from the screen.
- If possible, project the poster onto a wall before printing it to check formatting at actual size.
- Take a fine-line marker pen and white tape with you to the conference to fix any mistakes that you might have missed.
- Don't pin viewers down with an exhaustive tour of the poster. **K.P.**



Poster presenters should prepare different versions of their talk for passers-by and specialists.

an iPad at his poster session. Presenters can also direct viewers to large audio or video files, or to websites with additional data, using Quick Response barcodes that can be scanned by smartphones.

Purrington recommends that posters include laminated white space where the presenter can use a dry-erase marker to go through the research model. Get creative, he says, and have a section called ‘What I need help with,’ or overlap a panel with white tape so that extra information, such as a control experiment or an equation’s derivation, can be revealed. Images that don’t merit inclusion in the main poster can be attached on a ring.

Props can also serve as ice-breakers and attract viewers. “No one’s ever taken me up on this suggestion, but why not have a pitcher of beer and cups at your poster?” says Simon.

CONVERSATIONAL SCIENCE

Poster presenters should engage in conversations to help solve conundrums in data, get advice on improving the work or form possible future collaborations. At the ASCB poster session, Gutierrez seemed poised and in control. Her adviser at Rutgers and poster judge Alex Rodriguez, a cell biologist, who sat nearby, says that adopting an appropriate outlook in a poster session can be tricky for junior researchers. “Presenters have to be confident, but not defensive, and that’s a difficult line to ride, especially for young scientists,” he says. When a viewer challenged Gutierrez on her choice of cells, she pushed back in a friendly manner. “I hear what you are saying,” she responded, “but I still think this is a good model of wound healing.”

Perhaps most important, say judges and experienced presenters, is that presenters keep conversations with viewers friendly and two-way. It is a good idea to prepare

different versions of an oral presentation for casual passers-by, interested observers and interested specialists, says Faulkes. Presenters should practise their story, but not over-rehearse it. “You shouldn’t give it the same way twice,” he says.

Part of the preparation should be to anticipate tough questions. As a postdoctoral fellow, Lenart was caught off-guard when he presented research that used a proprietary compound provided by a pharmaceutical company. When people asked how the compound worked, he had to answer sheepishly that he could not say. Lenart advises presenters to prepare answers — or at least good excuses — for as many questions as they can think of. Presenters should do practice runs and discuss plans with a supervisor to decide what material is too sensitive to be shared.

Joseph Ramahi, the first-place winner of the ASCB poster competition and a doctoral student in cell biology at the University of California, Davis, thinks that his enthusiasm for presenting posters probably had an impact on the judges. “The questions I’ve been asked from giving posters have helped my project to grow,” he says. “Smile, and thank visitors for coming by, because they are helping you to get better, and that’s a really important part of science.”

Purrington believes that poster sessions are often under-rated as a place for scientists to air their “half-baked ideas”. The best part, he says, is picking other people’s brains about where a line of research should go next. “If you can get people excited enough about your question to want to be your colleague, or to give you a completely different take — that’s the fun of the poster session,” he says. “It can change your career.” ■

Kendall Powell is a freelance science writer based in Lafayette, Colorado.

UNITED STATES

Protect research ideas

A proposed US bill requiring all funded federal-grant applications to be posted on a government website would help competitors, including non-US scientists and businesses, to poach innovative research ideas, warns a group of US universities and scientific and professional societies. A letter sent on 15 February from the Coalition for National Science Funding (CNSF) in Washington DC to the US House of Representatives recommends that the bill require only abstracts to be posted. Samuel Rankin, associate executive director of the American Mathematical Society, a CNSF member, says that publicly open proposals would allow anyone to use researchers’ ideas in potential commercial applications, possibly even before the researcher can file a patent. “You’re giving away a lifetime of work,” says Rankin.

ENVIRONMENT

New observatory sites

The US National Ecological Observatory Network (NEON) will hire 15–20 ecologists who will collect ecological and climatic data, including soil, plant and animal samples, at three new observatory sites. NEON — a continent-wide network that will gather data over 30 years on the ecological impacts of climate change, invasive species and land-use changes — anticipates US\$60 million this year from the US National Science Foundation to build the observatories. The sites in Florida, Massachusetts and Colorado are expected to be completed by late 2013. NEON was established by the US National Science Board in 1999 to form a long-term network of ecological monitoring sites.

UNIVERSITIES

Top student city

Paris has been named ‘Best Student City 2012’ by QS Intelligence Unit, a company that compiles annual world university rankings. Paris earned the honour in part because of research institutions École Normale Supérieure Paris and École Polytechnique. London was second owing to the research institutions University College London and the London School of Hygiene and Tropical Medicine. Boston, Massachusetts, was third thanks to Harvard University and Massachusetts Institute of Technology, both in Massachusetts.